



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Essays on the selection and design of information structures

Sean Ferguson

Doctor of Philosophy



The University of Edinburgh

School of Economics

2020

Acknowledgements

I would like to thank my supervisors, Ina Taneva and József Sákovics, for the four years of advice, feedback, guidance and encouragement they've given me, and the many researchers and other PhD students at the University of Edinburgh School of Economics who gave helpful comments in seminars and other conversations.

The PhD would have been a much more gruelling process if I hadn't had a fantastic teaching experience, and thanks are due to everyone involved involved in the taught postgraduate courses at Edinburgh. In particular I'd like to thank Jonathan Thomas, Anna Babloyan, Sophie Bryan, Joe Stroud, Andrew Clausen, John Moore, and Paul Telemo.

I would also like to thank the Economic and Social Research Council, who supported this work.

Finally, I'd like to thank my wife Chloë, without whom all of this would have been a lot less fun.

Abstract of Thesis

This thesis consists of three chapters, each of which is a self-contained analysis of a distinct theoretical setting. The chapters are linked by the overarching theme of choices about information structures. In each, I consider an agent or agents facing a choice about information provision, modelled as the selection of an information structure that determines the probability of a receiver observing particular messages, given particular states of the world being realised.

Chapter 1: Confirmation Bias and Markets for Information

If consumers of information suffer from confirmation bias, they may prefer a biased information source to a perfectly accurate one. This is a potential explanation for the existence of biased reporting in news media. I show how confirmation bias can be expressed in terms of utility derived directly from the relationship between a consumer's prior and posterior beliefs, and that this can induce preferences over signals. I derive the optimal signal for consumers, both when their preferences are entirely dictated by confirmation bias, and when an instrumental value of information is also present, showing that in each case consumers may prefer biased signals. When audience-maximising firms compete to offer information sources to consumers with these preferences, I show that the strategic environment is equivalent to a one-dimensional location game, and that results analogous to those in the existing literature on location games extend to this setting: in duopoly, the firms locate at the median of consumer preferences, but 'spread out' as the number of firms rises.

Chapter 2: How Informed should Voters be?

Ex ante identical agents participate in a simple majority vote on whether to adopt a new policy, of uncertain value to each agent. A utilitarian social planner chooses the information structure through which agents learn about their own IID, private valuations of the policy. With mandatory voting, perfect information is optimal if and only if a particular symmetry condition holds on the distribution of valuations. Otherwise, restricting information such that relatively-ambivalent voters sometimes make mistakes can increase total expected utility, by reducing the incidence of 'tyranny of the majority' outcomes. When indifferent voters are allowed to abstain, perfect information becomes non-optimal for symmetrical distributions as well, provided that the number of voters is sufficiently large.

Chapter 3: Information Sharing in Continuous Two-Player Games

Two *ex ante* identical agents with uncertain payoff types play a two-stage game. In the first stage, before learning their own payoff type, they select an information structure for their opponent; they may choose to reveal their payoff type to their opponent, conceal it, or partially reveal. Each player then draws one of two possible payoff types, observes their own type, and receives a message about the opponent's type generated by the structure their opponent chose in the first stage. In the second stage players choose actions from an interval. Payoffs depend only on the second stage actions (in which they are continuous)

and a player's own type. This framework encompasses versions of models in the literature on information sharing in oligopoly, extending them to allow for 'persuasive' information sharing rather than simply choosing revelation or concealment (or adding unbiased noise). I show how this clarifies existing results, by laying out the incentives to provide a more or less informative signal as marginal effects. This sheds some light on the known difference between Cournot and (differentiated) Bertrand competition when considering sharing of cost information (revelation in Cournot, concealment in Bertrand), which I show results from the fact that in Bertrand competition, there is a direct, cost-dependent impact of the opponent's action on a player's payoff that is not mediated through the player's own action. I also provide sufficient conditions for the existence of equilibrium in games of this type, and analyse some alternative second stage games, showing that there exist games and forms of payoff uncertainty for which multiple equilibria of the overall game exist, including equilibria with partial revelation, and despite equilibrium uniqueness being guaranteed for every second stage subgame.

Lay Summary

The papers that make up this thesis are all concerned with the factors determining choices about the provision of information. To be more specific, they are concerned with information *structures*, in the sense that these are choices about how people are informed which are made *before* the information is known to the decision-maker. The decision-maker is not discovering some information and then choosing whether to reveal it or distort it - they are ‘locking in’ choices about how precise and/or biased the information someone receives will be, without knowing for sure what the reality will turn out to be. Crucially, in all cases those who receive the information do understand the information structures themselves. They know if an information source is biased or inaccurate, and judge the information it provides accordingly. This puts constraints on what the decision-maker who chooses the information structure can achieve. To take an example from Chapter 3, say a firm is going to develop a production process for a new product and can choose whether to have a third party conduct that research publicly, revealing its cost of production to its competitors, or to conduct the research privately in-house. If the firm wants to convince its competitors its cost of production is low, and cannot convincingly do so by reporting its in-house research (because it is too easy to misrepresent, so the competitors would not trust it), choosing third-party research can do this, but only when the firm’s costs actually turn out to be low - the downside of the competitors being convinced in this case is that if the costs turn out to be high, they will know that and the firm will do less well in the market. This kind of trade-off is typical of the considerations a decision-maker faces when choosing between information structures - for a structure to be more convincing, it must also be more accurate.

Each of the chapters that follow considers what determines these choices in a particular setting - information for confirmation-biased consumers, information for voters, and information shared between decision-makers who interact with each other (like multiple firms facing the cost-information problem described above).

Chapter 1 considers the preferred information structures of people who are *confirmation-biased* - that is, they like to receive information that confirms and strengthens their existing beliefs, and dislike information that shows that their existing beliefs might have been wrong. I show that if they care mainly about whether information confirms or disconfirms their beliefs, over how strongly it does so (i.e. if a piece of information confirms their belief twice as strongly as another piece, they like it more, but not twice as much more), and are not either very confident or not very confident at all in their existing belief, then the information structures they most prefer will:

- be very accurate when their existing beliefs turn out to be true, always reporting the truth in that case
- be less accurate when their existing beliefs turn out to be false, sometimes reporting the truth, but sometimes reporting what they wanted to hear instead.

This makes ‘good news’ more likely than if the information structure was perfectly accurate all of the time, but it still allows them to get some amount of confirmation from the information they get - if the source reported what they wanted to hear all the time, then knowing that, their beliefs would not get any confirmation from the reports.

I also analyze the behaviour of media firms who compete to offer information sources to these kinds of consumers, when the consumers start out with varying initial beliefs about which of two possible outcomes is more likely. This means the consumers want different degrees of bias from their media sources. I show that when there are only two firms, they would both offer similar biases, around the median of what the different consumers would prefer. This could mean accurate reporting if there are an equal number of consumers with biases one way or the other, but if the majority of consumers start off skewed toward one possibility, then the median point that the firms choose will be a biased one. With more firms, the available choices will ‘spread out’ to cover the range of biases that different consumers demand.

Chapter 2 considers the optimal choice of information structures for voters in a two-option vote, when the voters have independent preferences. This means that even when well-informed, the voters may prefer different options, because they care about different things. In this case, voting as a decision-making method has problems with ‘tyranny of the majority’ outcomes. These outcomes involve the relatively-slight preferences of a majority overruling the strong preferences of a minority. If people could agree before finding out what their preferences are (say, before anyone knows what the likely impacts of some policy decision will be) to prevent these outcomes by allowing strength of preferences to be considered in choosing an outcome, they would, even if they are self-interested, because of the risk that they would end up being the ones in the harmed minority. However, this may not be possible, either because the majority vote decision-making system is determined by existing social norms and practices and hard to alter, and/or because it is not possible to determine the strengths of preferences, in particular because people have an incentive to exaggerate when that information is taken into account in decision-making.

I show that in these cases, providing perfect information to voters that allows them to determine precisely which outcome they would prefer may actually be harmful, by increasing the impact of ‘tyranny of the majority’ outcomes. It may instead be optimal to provide enough information to ensure that those who have strong preferences one way or the other are aware of it and vote for their preference, but to allow those who are relatively-ambivalent (i.e. those who would need very fine-grained and precise information to determine precisely which outcome they prefer) to make ‘mistakes’, because this can reduce the risk of their determining the outcome when it causes significant harm to those with stronger preferences. This is particularly the case when voting is not mandatory and ambivalent voters can abstain from voting, in which case it is typically optimal to let the most ambivalent voters abstain rather than devote resources to providing the precise information that would cause them to choose one way or the other based on a small degree of preference.

Chapter 3 considers situations in which two agents (people or firms) who are going to strategically interact in some way - competing in a market, for example - can choose whether the

other agent can gain information about their own motivations and payoffs. They make this decision before they learn that information themselves, and before the strategic interaction takes place. One example of this is the situation I described at the beginning of this summary, in which firms decide whether their competitor will also be informed about their costs.

I set out a class of strategic interactions (games) for which I prove that an equilibrium (mutually-optimal solutions for all agents) of the information-sharing problem exists, and analyze how that equilibrium depends on the kind of strategic interaction the agents are going to have. I show that incentives to share or conceal information depend on two effects - how changing the information the other agent gets will change what they do, and how the agent making the decision's likely preferences about what the other agent does change if the information the other agent is getting changes. Remember that with information structures, the only way to make information suggesting that the first agent has some particular preferences more convincing for the other agent is to make it more likely that the first agent actually *does* have those preferences when that information is sent. That means it matters whether the first agent benefits more or less from convincing their opponent when they actually do have those preferences than when they have other possible preferences.

This helps to explain existing results about the example with firms sharing cost information. If firms produce substitute (competing) goods and compete on quantity, treating the market price as fixed, then they always want to convince competitors that they have low costs. However, the benefit they get from this is higher when they do actually have low costs, because when they produce more goods, they get more benefit from a competitor reducing its production level. This leads firms to want to reveal cost information - it is worth revealing that their costs are high when that occurs, because the benefit to revealing that their costs are low when that occurs is relatively large. On the other hand, when firms compete on price, they want competitors to believe that their costs are high, because the competitors then expect a higher price from the firm and raise their own prices accordingly, leading to more demand for the firm's product. However, this benefit is smaller when the firm's cost is actually high, because then they charge a higher price and produce less, so they benefit less from their competitors raising prices. This leads firms to want to conceal cost information.

In addition to these cases, I also consider some other types of uncertainty, and show that as well as fully concealing or fully revealing information, it is possible for agents to choose partially-revealing information structures that provide some information but not perfect accuracy. This does not happen with information about costs as described above, but it could happen if the information concerned which 'direction' the agent would like their opponent to change their behaviour in - for example, whether a firm would benefit from a competitor increasing its production or lowering it.

Contents

Acknowledgements	i
Abstract of Thesis	ii
Lay Summary	iv
1 Confirmation Bias and Markets for Information	1
1.1 Introduction	1
1.2 The Basic Model	3
1.2.1 Psychological Preferences and Confirmation Bias	3
1.2.2 Consumer Payoffs	5
1.2.3 Firms, Signals, and Timing	6
1.3 Pure Confirmation Bias	8
1.3.1 Consumer Preferences and Optimal Signals	8
1.3.2 The Distribution of Preferences	12
1.3.3 Duopoly	15
1.3.4 More Than Two Firms	17
1.4 Partially-biased Preferences	20
1.4.1 Consumer Preferences	20
1.4.2 Firms	23

1.5	Conclusion	24
	Bibliography	25
	Appendix	27
2	How Informed Should Voters Be?	48
2.1	Introduction	48
2.2	Baseline Model with Mandatory Voting	51
2.2.1	Structure and Timing	51
2.2.2	The Planner's Problem	53
2.2.3	Optimal Information Structures	57
2.2.4	Alternative majority thresholds	61
2.3	Abstention	62
2.4	Conclusion	69
3	Information Sharing in Continuous Two-Player Games	71
3.1	Introduction	71
3.1.1	Related Literature	75
3.2	An example: Cournot with cost uncertainty	77
3.3	General Model	80
3.3.1	Second-stage games with belief-continuous equilibria	83
3.3.2	First-stage best responses and the ex-ante value of posteriors	85
3.4	Quadratic games	88
3.4.1	Externalities	93
3.4.2	Cross-Effect Uncertainty	97

Chapter 1

Confirmation Bias and Markets for Information

1.1 Introduction

Since the mid-2000s, a variety of authors have constructed models in which information providers (in particular, media firms) might bias or ‘slant’ the information they provide. The discussion has been driven by public suspicion of media reporting in the United States; Pew Research Center (2017) provides evidence of widespread belief in media bias among the US public, at least as regards politics. Evidencing actual bias, rather than the belief in it, is difficult without relying on a common premise as to ‘the facts’ of any particular event, but it is easier to support the claim that the bias of different information providers differs in direction or degree. In that case, one can simply demonstrate that different sources report differently, without having to take a position¹. Groseclose and Milyo (2005) attempt to do this quantitatively, based on the sources different outlets cite. They find considerable variation across different US media outlets. Gentzkow and Shapiro (2006) take the simpler approach of demonstrating variation by quoting notably different reports of the same events from different outlets. To step away from the US for a moment, Smith (2017) demonstrates that this picture of varying political biases matches public perceptions of the UK newspaper market. Here beliefs about the political slants of the various newspapers were not strongly dependent on the beliefs of the respondents themselves; consumers of different political affiliations might differ in their assessments of the *degree* of bias of different sources, but there was consensus on the *direction* (i.e. which papers lean left and which right).

¹The fact of different sources reporting differently is not *conclusive* evidence of the existence of bias. Different sources could have access to different information, leading to varying reports without deliberate bias. Given the existence of news agencies and the informational resources available to modern media organisations, this is not a compelling explanation for variation in reporting, although there are particular circumstances in which it is more likely to apply: breaking news, for example.

While an economist’s first instinct might be to argue that market competition should eliminate biased reporting (Mullainathan and Shleifer (2005) call this the ‘traditional conception of the demand for news’), there has been no shortage of potential explanations for how bias might originate and persist. In this paper, I develop a model in which consumers have a preference for biased information sources, which, naturally enough, can lead firms to provide them. The source of this preference is *confirmation bias* on the part of the consumers. This psychological concept is something of an umbrella term for a collection of phenomena that all involve some kind of bias towards maintaining pre-existing beliefs; Nickerson (1998) provides a breakdown, along with details of empirical evidence for each of the different forms of confirmation bias he identifies. I am concerned with what Nickerson refers to as ‘Hypothesis-determined Information Seeking’, meaning a tendency to seek out information that confirms beliefs already held and avoid information that contradicts them².

This relates naturally to the concept of *selective exposure* from Communication Studies, meaning a tendency to choose to consume information matching pre-existing beliefs, and/or to avoid information contradictory to them. Historically, assessments of the empirical evidence for selective exposure have varied considerably, but recent work has provided evidence of it specifically in relation to political information. Knobloch-Westerwick and Meng (2009) find evidence of selective exposure in a laboratory setting in which subjects could read opposing articles on four political issues (gun ownership, abortion, health care regulation, and minimum wages), and Stroud (2008) uses survey data to argue for dual causality in media consumption and political beliefs, consistent with selective exposure: individuals consume media in accordance with their beliefs, and the media they consume impacts how their beliefs change over time.

The concepts of confirmation bias and selective exposure may seem somewhat identical at this point, and there is certainly some overlap in the way they have been used in the literature. I introduce both terms because I want to make a distinction between a preference for information that confirms one’s beliefs (confirmation bias) and a preference for information *sources* that are more likely to provide such information (selective exposure). Given these definitions, confirmation bias can be expected to lead to selective exposure, but selective exposure does not necessarily imply the existence of confirmation bias.

To see why, note that if information sources differ in quality (the accuracy of their reports) and a rational consumer cannot directly observe the quality of a source, they will form beliefs about that quality based on how closely the reports match their prior beliefs. If time or cognitive resources are limited, this could lead consumers to engage in selective exposure, even though they would simply opt for the most accurate information *if the quality of information sources was known*. The reputation model of Gentzkow and Shapiro (2006) demonstrates how this kind of setting could induce media bias.

The model I set out in this paper takes the accuracy and biases of the information structures firms choose to offer as public knowledge. Given this I demonstrate how, when consumers

²The model of Rabin and Schrag (1999) provides an example of an alternative form of confirmation bias, in which agents interpret the information they receive as more supportive of prior beliefs than it actually is.

are affected by confirmation bias, they can prefer (and firms choose to offer) biased information. The resulting location game between firms competing for consumers with heterogeneous preferences is similar to the model of media markets in Mullainathan and Shleifer (2005). They also model consumers of information as motivated by confirmation bias, but they define preferences directly over reports, whereas I have drawn on recent work on psychological information preferences (such as Lipnowski and Mathevet (2018)) to model confirmation bias as a preference over future beliefs. In Mullainathan and Shleifer’s model, firms’ distortions of information are linear, and reversible by a rational consumer. It is unclear what effect the biased information would have on the beliefs of the consumers. In my model, consumers are rational Bayesians who understand the information structures they are choosing between. When they prefer a biased information structure, they gain a lower probability of being told that they are wrong, but there is a tradeoff: the confirmatory reports are less convincing, reducing their payoff. This paper shows how an acceptance of this tradeoff could lead consumers to prefer biased information structures, in a framework that gives a full account of their beliefs and could provide the basis for a model of how consumer beliefs and media biases develop together over time.

Section 2 introduces the model of consumer preferences and the structure of the firms’ game. In **Section 3** I consider consumers whose preferences over information sources are purely determined by confirmation bias, derive their optimal signals and show how the firms’ game, which is two-dimensional, can be reduced to a one-dimensional location game. **Section 4** extends the model to include an instrumental value of information, and demonstrates that equivalent results to those in **Section 3** still hold. **Section 5** concludes.

1.2 The Basic Model

1.2.1 Psychological Preferences and Confirmation Bias

I model confirmation bias as a form of *psychological preference*, in which a consumer’s posterior beliefs are a fundamental component of the utility she attaches to particular outcomes. Various psychological phenomena can be modelled in this way, and Lipnowski and Mathevet (2018) provide several examples. The appropriate form needed for such a beliefs-based utility function to represent the phenomenon of confirmation bias merits some discussion.

Lipnowski and Mathevet’s examples include a formulation that seems aimed at confirmation bias, or at least a related class of phenomena. They refer to it as ‘prior-bias’ or ‘stubbornness’, and give the following utility function³:

$$U(a, \omega, \mu'; \mu) = i(a, \omega) - \rho |\mu' - \mu|$$

Here ω is the state of the world, which can take on one of two values, μ is the prior belief

³For clarity, I use the same notation as the model I present later in the paper, rather than Lipnowski and Mathevet’s original notation.

of the agent expressed as the subjective probability she places on one of the two possible states, μ' is her posterior belief after receiving a message carrying some information, and a is an action she takes having formed that posterior. Note that the measure on $\mu' - \mu$ is just the absolute value, since the state is binary. The first term, $i(a, \omega)$, is dependent only on the relationship between her action and the state. The second term, which is weighted by the positive parameter ρ , is the psychological element of the agent's utility: she incurs disutility based on the absolute difference between her prior and posterior beliefs.

This is certainly something like confirmation bias; the second term represents a desire to cling to one's prior belief. However, there is no 'confirmation' for the agent to seek here. Any information whatsoever will change her beliefs and incur psychological disutility. This captures information-aversion, not confirmation bias. They differ because confirmation bias is not an attachment on the agent's part to the specific probabilities she attaches to each possible state, but to the *state she considers more likely*; that is, to her expected state.

To make this clearer, consider a more concrete example. Say the binary state reflects the success or failure of some politically-charged policy; for example, whether a controversial healthcare reform package has succeeded in reducing healthcare costs. If our hypothetical confirmation-biased consumer starts out as a supporter of the policy, attaching a high probability (say 0.7) to its success, we do not expect her to be averse to information that strengthens that belief; it is the belief that the policy has probably succeeded that she is attached to, not the specific odds of 70%. The psychological utility should be maximised at a posterior of 1, not 0.7.

The natural first step towards fixing this problem is to simply remove the absolute value operator from the second term, and then determine that term's sign based on whether the prior μ is greater or less than $\frac{1}{2}$:

$$u(a, \omega, \mu'; \mu) = \begin{cases} i(a, \omega) + \rho(\mu' - \mu) & \mu > \frac{1}{2} \\ i(a, \omega) - \rho(\mu' - \mu) & \mu < \frac{1}{2} \end{cases}$$

This captures confirmation-biased preferences over *outcomes* well enough; the psychological payoff increases when the information received increases the agent's confidence in the state she originally held to be more likely, and decreases when it does the opposite. However, it still leaves us with significant problems:

- An ambivalent agent who does not hold either state to be more likely ($\mu = \frac{1}{2}$) should not be influenced by confirmation bias, and payoffs should not be discontinuous at $\mu = \frac{1}{2}$, as it is rather implausible that moving from a very slight belief in favour of one state to a very slight belief in the other direction would cause a dramatic reversal in preferences.
- While the utility function does capture psychological preferences over outcomes, for a Bayes-rational agent it does not affect preferences over information sources. An information source ('signal') can be equated to a distribution over posterior beliefs with expectation equal to the prior belief, as set out in Kamenica and Gentzkow (2011).

Since the psychological payoff above is linear in the posterior μ' and symmetrical around the prior μ , any information source will result in the same expected psychological payoff (of 0).

A utility function that aims to plausibly represent confirmation bias in preferences over information sources must address both of these issues. I formally present such a utility function in the next subsection.

1.2.2 Consumer Payoffs

There is a binary state of the world, $\omega \in \{A, B\}$. A consumer is born with some prior belief about the state, expressed as $P(A) = \mu$. On receiving information, she updates her beliefs using Bayes' rule and forms a posterior, $P(A) = \mu'$. The utility U she receives from a given outcome depends entirely on the true state ω , her prior belief μ , and her posterior belief μ' , and is given by a function of the following form:

$$U(\mu', \omega; \mu) = i(\mu', \omega) + \rho u(\mu'; \mu)$$

Here i represents the instrumental value of information. It only depends on the posterior and the true state. Typically it will be useful to think of this payoff as resulting from the consumer's choice of an optimal action given her beliefs, where the payoffs to possible actions depend on the true state. The i function must be weakly increasing in the accuracy of the posterior belief; that is:

$$\frac{\partial i}{\partial \mu'} \begin{cases} \geq 0 & \omega = A \\ \leq 0 & \omega = B \end{cases}$$

The second component of the utility function, u , represents the psychological preferences associated with confirmation bias. It depends on the relationship between the prior μ and the posterior μ' . I formulate it as follows:

$$u(\mu'; \mu) = \begin{cases} (\mu - \frac{1}{2})(\mu' - \mu)^{\frac{1}{2}} & \mu' \geq \mu \\ -(\mu - \frac{1}{2})(\mu - \mu')^{\frac{1}{2}} & \mu' < \mu \end{cases}$$

Weighting by $(\mu - \frac{1}{2})$ eliminates the problem at and around $\mu = \frac{1}{2}$ discussed in the previous section. This function is continuous at $\mu = \frac{1}{2}$, and the weighting captures the plausible assumption that the degree of confirmation bias depends on the strength of the original prior, approaching 0 as the prior approaches pure ambivalence at $\frac{1}{2}$. Since preferences reverse at $\mu = \frac{1}{2}$, in the sense that consumers with $\mu > \frac{1}{2}$ have a psychological payoff increasing in the posterior where the psychological payoff of consumers with $\mu < \frac{1}{2}$ is decreasing in the posterior, it is important to distinguish the two cases. I will refer to consumers with $\mu > \frac{1}{2}$ as **A-biased consumers** and to consumers with $\mu < \frac{1}{2}$ as **B-biased consumers**.

Giving the change in beliefs an exponent between 0 and 1 (I use $\frac{1}{2}$ as a tractable example) addresses the linearity problem, essentially making the consumer risk-averse with regard to

her confirmation bias; she gets decreasing marginal utility from a greater degree of confirmation. Essentially, the fact of (dis)confirmation is more important to her than the degree of it. Figure 1.1 graphs the psychological payoff as a function of the posterior for a consumer with a prior of 0.6; note how the marginal impact of an increased change in beliefs decreases as she moves further from her prior. If the exponent on the change were instead greater than 1, this would lead the consumer to engage in risk-seeking behavior; among signals with similar overall accuracy, she would prefer those biased *against* her preferred state, because the lower probability of confirmation would be compensated by a higher degree of certainty when it did occur. It is not entirely implausible that such individuals exist, but it is not generally consistent with the phenomenon of selective exposure, so I will focus on the risk-averse case.

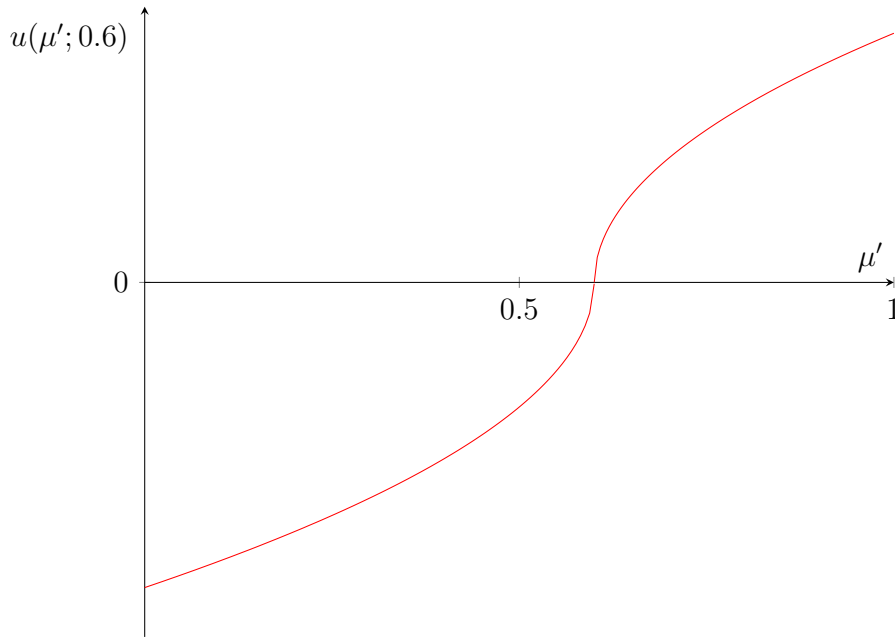


Figure 1.1: Psychological payoffs to varying posterior beliefs, for $\mu = 0.6$

1.2.3 Firms, Signals, and Timing

N firms face a unit mass of consumers who have heterogeneous prior beliefs, distributed according to a commonly known cumulative distribution function $F(\mu)$. Firms offer signals to consumers; each signal is an information structure which will generate a single message. The message is drawn from the set $\{a, b\}$, with probabilities conditional on the state given by $\alpha = P(a|A)$ and $\beta = P(b|B)$. A firm's choice of signal is the choice of values for these two probabilities, subject to the labelling restriction $\alpha \geq 1 - \beta$. This restriction ensures that the message a is a report that the state is A , in the sense that it is more likely to be sent in state A than in state B (and ensures the analogous relationship between message b and state B). A signal that violates this restriction can be converted into one that satisfies it by simply relabelling the two messages.

Note that the firm's decision amounts to choosing the accuracy of its reporting in each possible state. The labelling restriction rules out consistent, across-all-states misreporting: a firm cannot send message a whenever the state is B and message b whenever the state is A , since this is informationally equivalent to just telling the truth. It does, however, allow for biased reporting. A firm can offer a signal that is more accurate in one state than in the other, all the way up to a signal that sends the same message every time (for example, $\alpha = 1$, $\beta = 0$, which always sends message a). I refer to signals as **A-biased signals** if they are more accurate in state A ($\alpha > \beta$) and **B-biased signals** if they are more accurate in state B ($\beta > \alpha$). Figure 2 depicts the possible signals graphically.

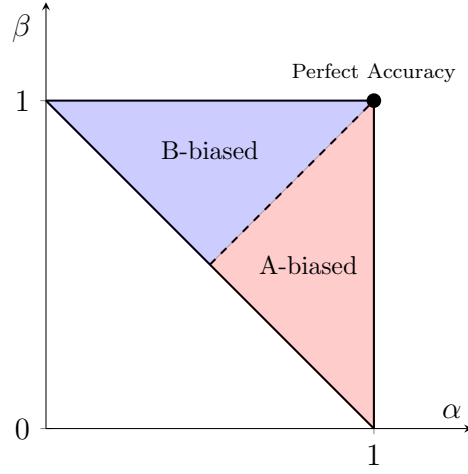


Figure 1.2: The space of possible signals

Each firm aims to maximise its market share; that is, the proportion of consumers who choose that firm's signal. I am concerned here with a scenario in which advertising is their source of revenue, and signals are free to consumers, as is typical of much modern digital media. I denote the signal offered by a representative firm i by $S_i = (\alpha_i, \beta_i)$. The set of all possible signals is \bar{S} , where:

$$\bar{S} = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha, \beta \in [0, 1], \alpha \geq 1 - \beta\}$$

I allow mixed strategies, so a strategy for a firm is an element of $\Delta(\bar{S})$.

The timing of the game is as follows:

- Firms simultaneously announce their signals.
- Each consumer selects the signal that maximises her expected utility. If multiple signals provide the highest available expected utility, the consumer selects from them randomly, with equal probability.
- Each signal generates a message, which is sent to the consumers who have chosen that signal. They update their beliefs using Bayes' rule, potentially take an action based on their updated beliefs, and receive utility according to the utility function given in Section 2.2.

An equilibrium of this game is a strategy for each firm that maximises the firm's expected market share given the strategies of the other firms and the fact that consumers maximise expected utility.

1.3 Pure Confirmation Bias

I begin by analysing the simplest possible case, in which $i(\mu', \omega) = c$ for some constant c , so that the utility function is:

$$U(\mu', \omega; \mu) = c + \rho u(\mu'; \mu)$$

I will refer to preferences represented by the above utility function as **pure confirmation bias**.

Pure confirmation bias implies that consumers get some fixed payoff from consuming a signal, independent of its informativeness (perhaps some entertainment value derived from reading the news), and otherwise only care about their psychological payoff. In general, it is not particularly plausible that confirmation bias would be the only driver of a consumer's choice of information source, but addressing this case will help to clarify the effect of the psychological preferences, and provide a baseline from which to extend to a more general case.

1.3.1 Consumer Preferences and Optimal Signals

The properties of a given signal (the values of α and β) enter into a consumer's expected psychological utility in two ways; they determine the probabilities she attaches to the two possible outcomes, message a and message b , and they determine the posterior beliefs she will update to (and consequently her psychological payoff) given either outcome. Denote the probability a consumer with prior μ attaches to receiving a message m from a given signal S by $P(m|S, \mu)$. These probabilities are given by:

$$P(a|S, \mu) = 1 - P(b|S, \mu) = \mu\alpha + (1 - \mu)(1 - \beta) = \mu(\alpha + \beta - 1) + 1 - \beta$$

Denote the posterior belief generated by receipt of a message m , given a signal S , as $\mu'(m|S, \mu)$. These posteriors are given by Bayes' rule as:

$$\mu'(a|S, \mu) = \frac{\mu\alpha}{P(a|S, \mu)}$$

$$\mu'(b|S, \mu) = \frac{\mu(1 - \alpha)}{P(b|S, \mu)}$$

Using these and taking expectations of the consumer's utility function gives the expected utility of a given signal:

$$E(U|S) = \rho(\mu - \tfrac{1}{2}) \left[(\mu'(a|S, \mu) - \mu)^{\frac{1}{2}} P(a|S, \mu) - (\mu - \mu'(b|S, \mu))^{\frac{1}{2}} P(b|S, \mu) \right] + c$$

Substituting in the expressions for the two possible posterior beliefs and factoring out common terms gives the following:

$$E(U|S) = \rho(\mu - \tfrac{1}{2}) \mu^{\frac{1}{2}} (1 - \mu)^{\frac{1}{2}} (\alpha + \beta - 1)^{\frac{1}{2}} \left[P(a|S, \mu)^{\frac{1}{2}} - P(b|S, \mu)^{\frac{1}{2}} \right] + c$$

Note that for a given μ , $\rho(\mu - \frac{1}{2}) \mu^{\frac{1}{2}} (1 - \mu)^{\frac{1}{2}}$ is constant and does not vary with the properties of the signal. Therefore, when the only consideration is confirmation bias, a consumer's preferences over signals can be fully captured by the following expression for the value of a signal:

$$V_{pure}(S; \mu) = \begin{cases} (\alpha + \beta - 1)^{\frac{1}{2}} [P(a|S, \mu)^{\frac{1}{2}} - P(b|S, \mu)^{\frac{1}{2}}] & \mu > \frac{1}{2} \\ (\alpha + \beta - 1)^{\frac{1}{2}} [P(b|S, \mu)^{\frac{1}{2}} - P(a|S, \mu)^{\frac{1}{2}}] & \mu < \frac{1}{2} \\ 0 & \mu = \frac{1}{2} \end{cases}$$

Note that the first term in this expression, $(\alpha + \beta - 1)^{\frac{1}{2}}$, is a measure of the informativeness of the signal independent of the relative probabilities of the states, while the second term, $P(a|S, \mu)^{\frac{1}{2}} - P(b|S, \mu)^{\frac{1}{2}}$ for $\mu > \frac{1}{2}$, represents the relative probability of good news compared to bad news. The trade-off the consumer faces is laid out here, as increasing the second term to its maximum value of 1 (a signal with $\alpha = 1$ and $\beta = 0$, which always reports good news) would reduce the first term to 0, since the signal would be entirely uninformative.

In discussing the resulting preferences over signals, I will often focus on the first case above, **A-biased** consumers with $\mu > \frac{1}{2}$. Because of the symmetry of the setup, analogous results (substituting $(1 - \mu)$ for μ and swapping α and β) hold for **B-biased** consumers.

Lemma 1. *For $\mu > \frac{1}{2}$, if $\alpha \geq \beta$, $V_{pure}(S; \mu)$ is increasing in α .*

Proof. $P(a|S, \mu)$ is increasing in α ($\frac{dP(a)}{d\alpha} = \mu$), so $P(a|S, \mu)^{\frac{1}{2}} - P(b|S, \mu)^{\frac{1}{2}}$ is increasing in α . Since $\alpha \geq \beta$ and $\alpha \geq 1 - \beta$, $\alpha \geq \frac{1}{2}$. Combining this with $\mu > \frac{1}{2}$ gives:

$$P(a|S, \mu) = \mu\alpha + (1 - \mu)(1 - \beta) \geq \mu\alpha + (1 - \mu)(1 - \alpha) > \frac{1}{2}$$

This implies that $P(a|S, \mu)^{\frac{1}{2}} - P(b|S, \mu)^{\frac{1}{2}} > 0$. Since $V_{pure}(S; \mu)$ is the product of two terms that are both positive and increasing in α , it is increasing in α . \square

Lemma 1 shows that among A-biased signals, an A-biased consumer always prefers a signal to generate more accurate reports in state A, for any given accuracy in state B.

Lemma 2. For $\mu > \frac{1}{2}$ and signals S_x and S_y with $\alpha_x + \beta_x = \alpha_y + \beta_y$ and $\alpha_x > \alpha_y$, $V_{pure}(S_x; \mu) > V_{pure}(S_y; \mu)$.

Proof. $P(a|S, \mu)$ is increasing in α and decreasing in β , and the reverse is true for $P(b|S, \mu)$. Therefore $(P(a|S, \mu)^{\frac{1}{2}} - P(b|S, \mu)^{\frac{1}{2}})$ is increasing in α and decreasing in β . Since $\alpha_x > \alpha_y$ and $\beta_x < \beta_y$, it follows that:

$$P(a|S_x, \mu)^{\frac{1}{2}} - P(b|S_x, \mu)^{\frac{1}{2}} > P(a|S_y, \mu)^{\frac{1}{2}} - P(b|S_y, \mu)^{\frac{1}{2}}$$

As $(\alpha_x + \beta_x - 1)^{\frac{1}{2}} = (\alpha_y + \beta_y - 1)^{\frac{1}{2}}$, $V_{pure}(S_x; \mu) > V_{pure}(S_y; \mu)$. □

Lemma 2 implies that for a given ‘total accuracy’, an A-biased consumer would prefer that as much of that accuracy as possible be assigned to state A.

Lemma 3. For an A-biased consumer with pure confirmation bias, any optimal signal has $\alpha = 1$.

Proof. Assume an optimal signal S_x exists with $\alpha_x < 1$. Construct a new signal S_y with $\alpha_y = 1$, $\beta_y = \beta_x + \alpha_x - 1$. By Lemma 2, this new signal is preferred to the original signal, which therefore cannot be optimal. □

This result is intuitive, since we would expect a biased consumer to prefer accuracy in the event that her biases turn out to be correct. To return to the healthcare-policy example I introduced in Section 1.2.1, let us say that state A represents the success of the policy, and state B the failure of the policy. If a consumer suffers from confirmation bias and has a prior belief that the policy is likely to succeed, then if the policy does succeed, she would prefer her ideal news source to report the fact accurately. However, since the consumer is Bayesian, accuracy in the preferred state does come at a cost. The fact that the optimal signal has $\alpha = 1$ implies that any b report from such a signal produces certainty that the state is B ; the worst possible outcome. This provides an incentive to seek a reduced probability of such reports (a lower β), which must be weighed against the fact that this will make a reports less convincing, and therefore less rewarding.

Lemma 4. For $\mu > \frac{1}{2}$, $V_{pure}((1, \beta); \mu)$ is strictly concave in β .

For a proof, see the appendix.

Lemmas 3 and 4 together imply that each A-biased consumer has a unique optimal signal, of the form $(1, \beta^*)$.

Proposition 1. For an A-biased consumer with pure confirmation bias, the unique optimal signal has $\alpha = 1$, $\beta = \beta_{bias}^*(\mu)$, where:

$$\beta_{bias}^*(\mu) = \min \left\{ \frac{2-\sqrt{2}}{4(1-\mu)}, 1 \right\}$$

For a proof, see the appendix.

Corollary 1. *For a B-biased consumer with pure confirmation bias, the unique optimal signal is $(\alpha_{bias}^*, 1)$ where:*

$$\alpha_{bias}^*(\mu) = \min \left\{ \frac{2-\sqrt{2}}{4\mu}, 1 \right\}$$

Proposition 1 demonstrates that a consumer's optimal signal can involve a degree of inaccuracy in the disliked state, *unless the consumer is confident enough to begin with* (either $\mu > \frac{2+\sqrt{2}}{4}$ or $\mu < \frac{2-\sqrt{2}}{4}$). If a consumer believes that the healthcare policy is likely to succeed, but lacks confidence in that belief, the worry that an accurate source will inform her that she was wrong may drive her to prefer a source that will sometimes report success even in the event of failure; but, crucially, not always, since in that case she could not derive any confirmation of her beliefs from the entirely predictable report of success.

It may seem counter-intuitive that the degree of bias that the consumer desires, as measured by $1 - \beta_{bias}^*(\mu)$, increases as the prior approaches $\frac{1}{2}$. After all, my reason for weighting the psychological utility by $\mu - \frac{1}{2}$ was to capture the idea that consumers with more ambivalent priors should suffer less from confirmation bias. This shows the limitations of the pure confirmation bias case. The consumers who are most confident in their priors would choose a perfectly accurate signal, because they believe an accurate signal has a high probability of reporting their preferred state, and they gain a larger increase in their belief in that case than they would with a biased signal. Less confident but still confirmation-biased consumers face a higher subjective risk of an unpleasant report, so they are willing to trade off a portion of the payoff from a pleasant report to reduce this risk. The fact that their confirmation-bias payoffs are smaller than those of the more confident consumers does not affect the outcome, since in this case there is no instrumental value to be derived from a more accurate signal.

Concavification

An alternative view of the optimal signals derived above comes from the concavification approach of Kamenica and Gentzkow (2011). They are concerned with the optimal information structure for a principal whose payoff varies with the posterior belief of an agent; here it is the agent's payoff that varies with the posterior, but the framework is the same. The best payoff she can attain can be found by constructing the concave envelope of her utility function (the smallest concave function that is everywhere larger than the original function) and evaluating it at her prior, since, for a Bayesian agent, a signal induces a distribution over posterior beliefs with expectation equal to the prior belief. Figure 1.3 shows how this applies in the current case, using the payoffs for $\mu = 0.6$ as in Figure 1.1 and showing the concave envelope. The optimal signal S^* allows the consumer to mix between a posterior of 0 and a posterior of $\mu'(a|S^*)$. The expected payoff of the optimal signal, $E(U|S^*)$ is given by the value of the concave envelope evaluated at the prior of 0.6⁴.

⁴The concavification argument demonstrates that the consumer-optimal signal I have derived is optimal among all possible signals; allowing more than two possible messages would not allow a consumer to do better.

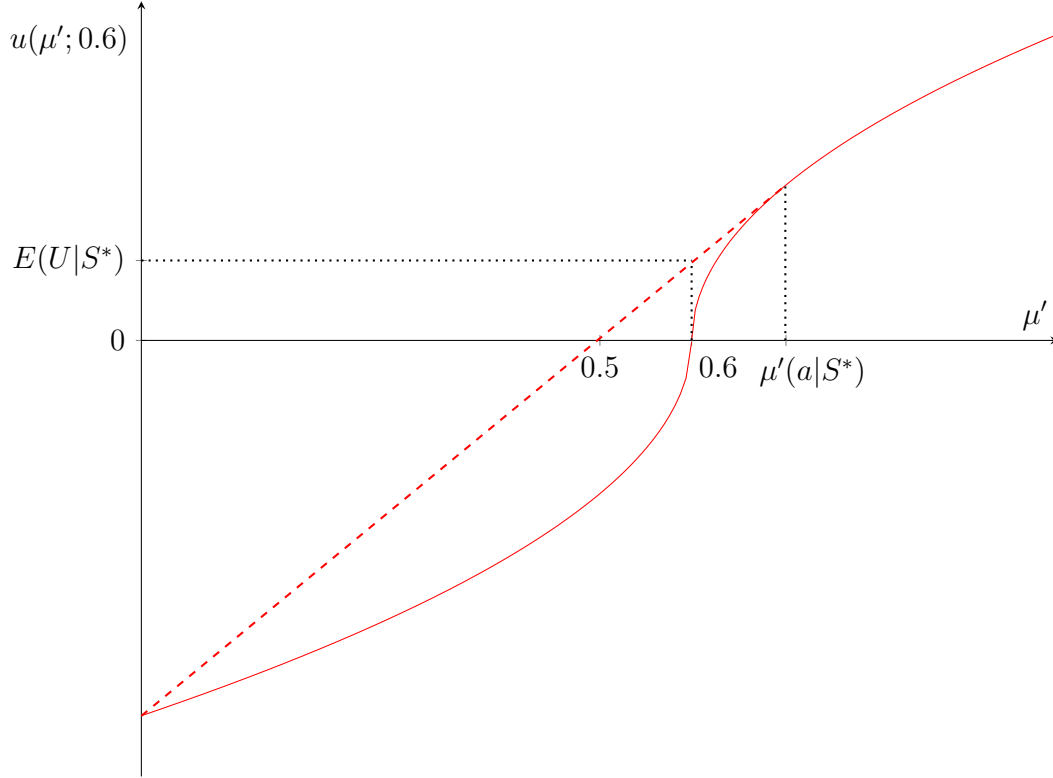


Figure 1.3: Concave envelope and optimal signal payoff for $\mu = 0.6$

1.3.2 The Distribution of Preferences

In order to understand the firms' problem, it is useful to consider how the distribution of priors (values of μ) in the consumer population determines the distribution of preferences over signals. We have seen that consumer-optimal signals must have either $\alpha = 1$, $\beta = 1$, or both. This means that despite the two-dimensional nature of signals, the set of consumer-optimal signals is one dimensional (it is the outer boundary of the triangle in Figure 2) and can be mapped onto an interval. Taking this to be $[0, 1]$, a signal $(\alpha, 1)$ or $(1, \beta)$ maps to a point $l \in [0, 1]$ as follows:

$$l(S) = \begin{cases} \frac{\alpha}{2} & \alpha \leq 1, \beta = 1 \\ 1 - \frac{\beta}{2} & \alpha = 1, \beta \leq 1 \end{cases}$$

Figure 1.4 shows how signals (labelled above the line) map to points on the interval (labelled below the line). The perfectly accurate signal (with α and β both equal to 1) is located at the centre of the interval, with A-biased signals to the right and B-biased signals to the left. Signals become more biased on moving further from the centre point (decreasing α by moving left from the centre, decreasing β by moving right from the centre); the endpoints equate to completely biased, entirely uninformative signals that always produce the same report.

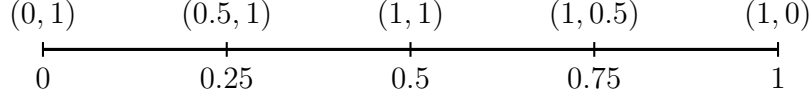


Figure 1.4: Signals mapped onto the $[0, 1]$ interval.

I now define a function, $l^* : [0, \frac{1}{2}) \cup (\frac{1}{2}, 1] \mapsto [0, 1]$, which maps from a confirmation-biased consumer's prior to their 'location' on the interval; i.e the point corresponding to their optimal signal.

$$l^*(\mu) = \begin{cases} \frac{2-\sqrt{2}}{8\mu} & 1 - \frac{2+\sqrt{2}}{4} \leq \mu < \frac{1}{2} \\ 2 - \frac{1-\sqrt{2}}{8(1-\mu)} & \frac{1}{2} < \mu \leq \frac{2+\sqrt{2}}{4} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

Assume that the distribution of consumer priors $F(\mu)$ is continuous at $\frac{1}{2}$. This is a fairly harmless assumption, since any mass of consumers in the population at $\mu = \frac{1}{2}$ will, in the pure confirmation-bias case, have no preference ordering over signals and can be assumed to choose randomly; the relevant population are consumers with confirmation bias, and nothing is lost here by treating $F(\mu)$ as the distribution of *their* priors. Given this, it is straightforward to convert the distribution of priors into the resulting distribution of locations on the interval, $G(l^*)$:

$$G(l^*) = \begin{cases} 0 & l^* < \frac{1}{2} - \frac{1}{2\sqrt{2}} \\ F(\frac{1}{2}) - F(\frac{2-\sqrt{2}}{4l^*}) & \frac{1}{2} - \frac{1}{2\sqrt{2}} \leq l^* < \frac{1}{2} \\ F(\frac{1}{2}) + 1 - F(1 - \frac{2-\sqrt{2}}{4(2-l^*)}) & \frac{1}{2} \leq l^* \leq \frac{1}{2} + \frac{1}{2\sqrt{2}} \\ 1 & l^* > \frac{1}{2} + \frac{1}{2\sqrt{2}} \end{cases}$$

I have defined these mappings on $[0, 1]$ for clarity, but the relevant interval is really the subset of $[0, 1]$ on which consumers are actually found; that is:

$$L = [\min \text{supp } G, \max \text{supp } G]$$

Note that if the distribution of priors, F , has full support on $[0, 1]$, then:

$$L = [\frac{1}{2} - \frac{1}{2\sqrt{2}}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$$

Define L_s as the set of signals that map into L . That is, any signal S such that $l(S)$ exists and $l(S) \in L$.

Arranging consumers on an interval in this way may seem a little abstract, but it will help clarify both consumers' preferences over signals and the character of the firms' problem. The

key intuition is that since consumers' preferred signals all fall within L_s , so should firms' equilibrium offerings. The only reason for firms to offer other signals (which would not be optimal for any consumer) would be to strike a balance between the preferences of consumers with differing priors. I next show that signals outside of L_s cannot do this any better than signals in L_s .

Lemma 5. *With pure confirmation bias, for any signal $S \notin L_s$, there exists a signal $S_0(S) \in L_s$ such that $S_0(S)$ is strictly preferred to S by every consumer with $\mu \neq \frac{1}{2}$.*

See the appendix for a proof. It is obviously the case that a given consumer will prefer some signal in L_s to a signal outside of that set, since L_s will contain her optimal signal, but this result is stronger: each signal outside of L_s is dominated by a *fixed* signal inside the set that is preferred by *all* consumers. This is an important result, because it implies that firms can never do better by offering signals not in L_s , and I can focus on equilibria in which all firms offer signals that map into L .

Note that Lemma 5 does not quite rule out equilibria in which firms offer such signals. For an arbitrary pair of signals, it is not guaranteed that a consumer indifferent between the two signals exists; even if the initial distribution of priors has full support on $[0, 1]$, it is possible for all A-biased consumers to strictly prefer one signal and all B-biased consumers the other. It follows that a firm could potentially offer a signal not in L_s without standing to gain customers by offering the universally-preferred signal implied by Lemma 5. Such a firm would be vulnerable to complete capture of its customer base by another firm, but if there are more than two firms, this would not necessarily be profitable for any other firm.

Such potential equilibria are of little interest. For any such equilibrium, there is a corresponding equilibrium in which every firm offers a signal in L_s ; simply replace the signals in their original strategies with the universally-preferred signals. Further, it does not make intuitive sense for firms to offer signals outside of L_s . For a firm to do so is to leave itself needlessly unprotected against unexpected moves by its opponents; simply put, there is no reason why a firm should ever offer a signal if another signal exists that is preferred by every consumer. To capture this intuition, I apply the 'trembling-hand-perfection' equilibrium refinement of Selten (1974).

Proposition 2. *With pure confirmation bias, any equilibrium in which any firm offers a signal not in L_s with positive probability is not trembling-hand-perfect.*

Proof. Consider a signal $S \notin L_s$. By Lemma 5, the strategy of offering S is weakly dominated by the strategy of offering $S_0(S)$: it never yields a lower market share, and yields a strictly higher market share if every other firm offers S . Therefore offering S cannot be an equilibrium strategy in any perturbed game, since dominance is strict when every other firm plays every signal (S and $S_0(S)$ in particular) with positive probability. Therefore the pure strategy of offering S cannot be part of a trembling-hand-perfect equilibrium.

The proof for mixed strategies is analogous: a strategy that puts a positive probability weight

on signals outside of L_s is weakly dominated by a strategy that shifts that weight onto the universally-preferred signals in L_s . \square

Having restricted attention to a one-dimensional space, my next result gives the role that distance plays within that space; a given consumer prefers the closest of all the signals located to her left, and the closest of all the signals located to her right.

Proposition 3. *With pure confirmation bias, if there exist two signals, $S_x, S_y \in L_s$, and there exists a consumer located at a point l^* , such that either $l^* \geq l(S_x) > l(S_y)$ or $l^* \leq l(S_x) < l(S_y)$, that consumer prefers S_x to S_y .*

See the appendix for a proof. Note that the spatial analogy here is imperfect, as distance only directly corresponds to the preference ordering among the signals in one direction: a consumer may prefer a signal to her left to one on her right (or vice versa) despite the preferred signal being ‘further away’ than the other.

Since there is a one-to-one mapping between signals in L_s and locations in the closed interval L , restricting attention to strategies in which firms only offer signals in L_s and applying Proposition 3 converts the analysis of the firms’ problem into a one-dimensional location game, similar to the classic ‘Main Street’ location game of Hotelling (1929) or the model of news markets in Mullainathan and Shleifer (2005).

1.3.3 Duopoly

When there are two firms, the trembling-hand-perfection refinement of Proposition 2 is not necessary; offering a signal outside L_s always provides the other firm with an opportunity to capture the entire market. Any signals offered in equilibrium must be in L_s . In fact, if there are no ‘gaps’ in the distribution of consumer priors (if F is strictly increasing), there is a unique equilibrium, as the next proposition demonstrates.

Proposition 4. *If F is strictly increasing, then there is a unique equilibrium in which both firms offer the signal corresponding to the median of G , the distribution of consumer preferences on L .*

Proof. Denote the two firms by x and y . If both firms offer the same signal, $S_x = S_y$, their expected market share is $\frac{1}{2}$. This implies that in any pure-strategy equilibrium, the expected market share of each firm must be $\frac{1}{2}$. If it was not, the firm receiving less than $\frac{1}{2}$ could do better by simply offering the same signal as the other firm.

If $S_x \notin L_s$, firm y can offer $S_0(S_x)$ and gain the entire market (by Lemma 5). Therefore signals outside of L_s cannot be offered in equilibrium and attention can be restricted without loss to signals with a location in L .

The situation is analogous to a standard two-firm Hotelling game without prices. Since F is strictly increasing, G is strictly increasing and has a unique median. Denote it by G_{med} . Assume that an equilibrium exists in which both firms receive an expected market share of $\frac{1}{2}$ and $l(S_x) \neq G_{med}$. Either $\lim_{l \downarrow l(S_x)} G(l) < \frac{1}{2}$ or $\lim_{l \uparrow l(S_x)} G(l) > \frac{1}{2}$. In the former case, firm y can gain a market share greater than $\frac{1}{2}$ by deviating to $l(S_x) + \epsilon$ for some sufficiently small ϵ ; by locating just to the right of firm x , they can gain a fraction, arbitrarily close to 1, of the consumers located to firm x 's right, who form a fraction of the total market equal to $1 - \lim_{l \downarrow l(S_x)} G(l) > \frac{1}{2}$. Therefore the firm can gain a market share greater than $\frac{1}{2}$. The same applies in the other case, using a deviation just to the left of firm x (to $l(S_x) - \epsilon$). Therefore this cannot be an equilibrium.

Finally, assume both firms offer signals located at G_{med} . Both firms have an expected market share of $\frac{1}{2}$, and clearly neither firm has any deviation that can increase their market share: any location to the left of G_{med} provides no access to the half of consumers with optimal signals to the right of G_{med} , and similarly signals to the right cannot gain consumers to the left. Therefore this is an equilibrium. \square

The assumption that F is strictly increasing is actually stronger than is necessary here; it guarantees that G has a unique median, which guarantees a unique equilibrium. Even if G does not have a unique median⁵, an analogous argument shows that a pair of strategies form an equilibrium if and only if there are no non-median points in their supports.

This result illustrates why mapping consumers onto an interval effectively characterises the strategic situation for firms. When a firm offers a signal, say an A-biased one, it divides consumers into two groups⁶: those A-biased consumers who would prefer that the signal be at least marginally more A-biased (consumers to the right of the firm's location), and consumers who would prefer the signal be at least marginally less A-biased (consumers to the left of the firm's location, including all B-biased consumers). A different firm offering an alternative signal can take an arbitrarily large fraction of one of these groups of consumers, by positioning marginally to the left or marginally to the right of the original firm, but cannot gain access to both groups⁷.

Note that for a distribution of priors that is symmetric around $\mu = \frac{1}{2}$, this leads both firms to offer a perfectly accurate signal. However, if the distribution is asymmetric or centred around another point, the firms may offer a biased signal. Further, the median of the pref-

⁵A median here is any point $l \in L$ with $G(l) = \frac{1}{2}$. A simple example of multiple medians is given by a discrete distribution with half of the consumers located at a point l_1 and half at a point $l_2 > l_1$. In this case any point l that satisfies $l_1 \leq l \leq l_2$ is a median.

⁶This is only strictly true if the signal is not located at a mass point of G . Otherwise there is a third group of consumers who prefer that signal to any alternative.

⁷The result notably differs from that of Mullainathan and Shleifer (2005), in which duopolistic firms offer extreme bias in order to differentiate themselves. This difference results from my focus on market share maximisation and advertising revenue, where Mullainathan and Shleifer's model has firms setting prices for their signals. This difference is the norm for one-dimensional location games with and without prices, and does not result from the differing formulations of consumer preferences.

erence distribution $G(l)$ does *not* generally correspond to the optimal signal of a consumer at the median of the distribution of priors, $F(\mu)$. In fact, they only correspond when the median of $G(l)$ is at $l = \frac{1}{2}$, the perfectly accurate signal. This is due to the fact that consumers at opposite ends of the distribution of priors (μ close to 0 or 1) can be close together in the preference distribution, and similarly, consumers who are close together in the prior distribution but on opposite sides of $\frac{1}{2}$ can be at opposite ends of the preference distributions. Consumers who have very different initial beliefs can have similar preferences over signals, because the combination of confirmation bias and confidence generates a preference for accuracy; consumers who are very confident that the state is A and consumers who are very confident that the state is B will both prefer an accurate signal, although they will expect different results from it.

1.3.4 More Than Two Firms

Duopoly is not typical of modern media markets. In the age of online media, a vast array of information providers are readily accessible. As described in the introduction to this paper, there is good reason to believe these providers offer varying biases. The natural question is whether this model reflects that differentiation when extended to more firms; is the clustering around the median of the preference distribution an artifact of the restriction to duopoly?

To answer this question in the affirmative, I begin by constructing an example equilibrium with three firms and a simple, discrete consumer distribution, in which the firms offer non-median signals with positive probability.

Example Three-Firm Equilibrium

Consumer Distribution

All consumers are one of three types, θ_1 , θ_2 , and θ_3 . Each type θ has an optimal signal located in L at a point $l^*(\theta)$, and $l^*(\theta_1) < l^*(\theta_2) < l^*(\theta_3)$. That is, θ_1 is the most A-biased type, θ_3 the most B-biased, and θ_2 falls in the middle. Further, θ_2 consumers have a prior that leaves them precisely indifferent between $l^*(\theta_1)$ and $l^*(\theta_3)$; if offered these two signals, they randomize evenly between them.

The distribution of types is as follows:

$$P(\theta) = \begin{cases} p & \theta = \theta_2 \\ \frac{1-p}{2} & \theta = \theta_1, \theta_3 \end{cases}$$

So this setting is doubly symmetric: the ‘centrist’ type is indifferent between the preferred signals of the two ‘extreme’ types, and the two ‘extreme’ types form equal fractions of the

total population.

Equilibrium Firm Strategies

When three firms face the consumer distribution described above, there is a symmetric mixed strategy equilibrium in which the firms offer all three types of signal, as long as the number of ‘centrist’ consumers is smaller than the number of either of the ‘extreme’ types; that is, as long as $p < \frac{1}{3}$.

To see this, assume that two firms play a strategy in which they offer a signal at $l^*(\theta_2)$ with probability q , and at $l^*(\theta_1)$ or at $l^*(\theta_3)$ with probability $\frac{1-q}{2}$. Let $\Gamma(l)$ denote the expected market share for a firm when it offers a signal at l . Then the third firm’s expected market share if it locates at one of the three consumer-optimal points are given by:

$$\Gamma(l^*(\theta_2)) = \frac{q^2}{3} + \frac{(1-q)^2 p}{2} + \frac{(1-q)^2(1+p)}{4} + \frac{q(1-q)(1+p)}{2}$$

$$\Gamma(l^*(\theta_1)) = \Gamma(l^*(\theta_3)) = \frac{(1-q)^2}{12} + \frac{(q(1-q) + q^2)(1-p)}{2} + \frac{(1-q)^2(1+p)}{4} + \frac{q(1-q)(1-p)}{4}$$

Any other location is weakly dominated by the consumer-optimal locations given that each of the other firms will locate at one of them, so equating these expressions and solving for q yields a symmetric-equilibrium strategy as long as a solution exists that falls between 0 and 1. Such a solution exists if and only if $p < \frac{1}{3}$, in which case the symmetric equilibrium strategy σ is given by:

$$\sigma(l) = \begin{cases} \frac{6p}{3p+1} & l = l^*(\theta_2) \\ \frac{1-3p}{2(3p+1)} & l = l^*(\theta_1), l^*(\theta_3) \\ 0 & \text{otherwise} \end{cases}$$

There is still an element of clustering here; if other firms stick to ‘extreme’ positions, the ‘centre’ becomes attractive as a firm can capture a large audience in the event that both other firms take the same ‘extreme’ position. Once there are enough centrists, $p \geq \frac{1}{3}$, this effect dominates and all firms offering $l^*(\theta_2)$ becomes the only equilibrium.

General Results

In general, specific equilibria are challenging to derive when there are more than two firms (and pure-strategy equilibria typically do not exist). However, where mixed-strategy equilibria exist, the following two propositions, which adapt results for more conventional location

games derived in Osborne and Pitchik (1986), demonstrate that as the number of firms rises, they will offer a wider variety of reporting biases catering to more 'extreme' consumers.

Proposition 5. *Given a consumer preference distribution G , for any $l \in \text{supp } G$ and any $\epsilon > 0$, there exists an M such that for any $N > M$, if there exists an equilibrium for N firms in which firm i 's strategy is given by H_i , then for at least one i , $\text{supp } H_i \cap [l - \epsilon, l + \epsilon] \neq \emptyset$.*

This proposition simply captures the intuitive conclusion that if there is a region (a range of biases) in which some positive number of consumers' optimal signal is located, then once there are sufficient firms in the market, it cannot be optimal for all firms to ignore that region. This follows straightforwardly from the continuity of consumer preferences in their prior; since consumers at l strictly prefer a signal at l to any other signal, and in particular any signal outside $[l - \epsilon, l + \epsilon]$, there is some interval around l in which consumers prefer a signal at l to any signal outside $[l - \epsilon, l + \epsilon]$. Since $l \in \text{supp } G$, there are a positive number of such consumers, who can be captured with certainty by locating at l if no firm ever locates in $[l - \epsilon, l + \epsilon]$. For a large enough number of firms this number must be above the lower bound on equilibrium payoffs⁸.

Note that for any interval that shares an endpoint with the support of G (a region of signals that stretches to the most 'extreme' biases in the consumer population), it is straightforward to give an upper bound on the number of firms for which it is possible for that interval to be neglected in equilibrium. Denote $\min \text{supp } G$ by g_{\min} , and consider an arbitrary location l , with $G(l) > 0$. Then if the number of firms n is large enough that $G(l) > \frac{1}{2n-1}$, firms must locate in the interval $[g_{\min}, l]$ with positive probability in any equilibrium.

I next adapt a key result of Osborne and Pitchik (1986) to the present setting. They prove that in one-dimensional location games with linear transportation costs, if symmetric mixed strategy equilibria converge as the number of firms rises, they must converge to the distribution of consumers.

Proposition 6. *If the distribution of priors F is continuous and there exists a subsequence $\{n_k\}$ of \mathbb{N} such that for each k , there exists a symmetric mixed strategy equilibrium in which each firm plays a strategy H_k , each H_k is twice continuously differentiable other than at $\frac{1}{2}$, and a twice continuously differentiable (other than at $\frac{1}{2}$) mixed strategy H exists such that $H_k \rightarrow H$, $H'_k \rightarrow H'$, and $H''_k \rightarrow H''$ uniformly, then $H = G$.*

See the appendix for a proof.

Propositions 5 and 6 suggest that as the number of firms grows, competition will drive firms to 'go where the consumers are' and offer a wider variety of biases.

⁸In any equilibrium, a firm always has a deviation to a strategy followed by some other firm, which in the worst case yields a payoff equal to half the other firm's equilibrium payoff. Combining this with the fact that payoffs must sum to one yields the condition that in an n -firm equilibrium, no firm can have an expected payoff lower than $\frac{1}{2n-1}$.

1.4 Partially-biased Preferences

I now turn to the more plausible case in which accurate information does have an instrumental value. Assume that each consumer, having consumed a signal and updated their beliefs, must take one of two actions, z_1 and z_2 , with payoffs as follows:

		State	
		A	B
Action	z_1	1	0
	z_2	0	1

Figure 1.5: State-matching Action Payoffs

Here the consumer is trying to correctly match her action to the state. In the healthcare-policy example, one could imagine that the consumer can change healthcare plan and that doing so will be beneficial if the policy is successful, or harmful if it is a failure.

Given that the consumer will choose her action to maximise her payoff given her posterior belief⁹, I can specify her instrumental payoff, $i_{match}(\mu', \omega)$:

$$i_{match}(\mu', \omega) = \begin{cases} 1 & \mu' \geq \frac{1}{2}, \omega = A \\ 1 & \mu' < \frac{1}{2}, \omega = B \\ 0 & otherwise \end{cases}$$

The full utility function in this case is:

$$U(\mu', \omega; \mu) = i_{match}(\mu', \omega) + \rho u(\mu'; \mu)$$

I will refer to preferences represented by this utility function as **partially-biased preferences**.

1.4.1 Consumer Preferences

Consider an A-biased consumer. Since μ' depends on the signal she consumes, the (*ex ante*) expected instrumental value given a signal $S = (\alpha, \beta)$ is a function of α and β :

$$E(i_{match}|\alpha, \beta) = \begin{cases} \alpha\mu + \beta(1 - \mu) & \mu'(b|\alpha, \beta) < \frac{1}{2} \\ \mu & \mu'(b|\alpha, \beta) \geq \frac{1}{2} \end{cases}$$

⁹I assume for simplicity that at $\mu' = \frac{1}{2}$ the consumer chooses z_1 .

An A-biased consumer's default action, based on her prior, is z_1 . Only a b report can change her action (an a report simply strengthens her confidence in her default action), and it only does so if it is sufficiently convincing ($\mu'(b|\alpha, \beta) < \frac{1}{2}$)¹⁰. All signals with insufficiently convincing b reports are equivalent in instrumental value; they vary in their effect on beliefs, but not on the resulting action. Note that despite its piecewise definition, the expected instrumental value is still continuous: substituting for $\mu'(b|\alpha, \beta)$ using Bayes' rule at $\mu'(b|\alpha, \beta) = \frac{1}{2}$ gives $\alpha\mu + \beta(1 - \mu) = \mu$, and μ' is continuous in α and β .

Substituting back into the full utility function gives the consumer's complete expected payoff as a function of the properties of the signal:

$$E(U|S, \mu) = E(i_{\text{match}}|\alpha, \beta) + \rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}}V_{\text{pure}}(S; \mu)$$

Similar results apply as in the case of pure confirmation bias. I focus once again on A-biased consumers.

Lemma 6. *For an A-biased consumer with partially-biased preferences, any optimal signal has $\alpha = 1$.*

Proof. Any signal that maximises $V_{\text{pure}}(S; \mu)$ has $\alpha = 1$, by Lemma 3. This is also true of any signal that maximises $\rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}}V_{\text{pure}}(S; \mu)$, since $\rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}}$ does not vary with the signal and is positive. $E(i_{\text{match}}|\alpha, \beta)$ is weakly increasing in α . Therefore, any signal that maximises $E(i_{\text{match}}|\alpha, \beta) + \rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}}V_{\text{pure}}(S; \mu)$ must have $\alpha = 1$. \square

Lemma 7. *For an A-biased consumer with partially-biased preferences, $E(U|(1, \beta))$ is strictly concave in β .*

Proof. $\rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}}V_{\text{pure}}((1, \beta); \mu)$ is strictly concave in β by Lemma 4. With $\alpha = 1$, $\mu'(b) = 0$, which gives $E(i_{\text{match}}|1, \beta) = \mu + \beta(1 - \mu)$, which is a linear in β . Therefore $E(U|(1, \beta))$ is strictly concave as it is the sum of a strictly concave function and a linear function. \square

Of course, these properties are familiar from Section 1.3.1, and the optimisation problem is quite similar to the pure case, as the following result makes clear.

Proposition 7. *For an A-biased consumer with partially-biased preferences, the unique consumer-optimal signal is $(1, \beta_{\text{match}}^*)$, where $\beta_{\text{match}}^* = \min\{\hat{\beta}(\mu, \rho), 1\}$ and:*

$$\hat{\beta}(\mu; \rho) = \frac{1}{2(1 - \mu)} \left[1 - \frac{\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) - 1}{(2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{\frac{1}{2}}} \right]$$

Corollary 2. *For an A-biased consumer with partially-biased preferences, $\hat{\beta}(\mu; \rho)$ is decreasing in ρ , and β_{match}^* approaches β_{bias}^* as ρ goes to infinity.*

¹⁰Recall that $\mu'(b|\alpha, \beta)$ is the posterior belief following a b report from a signal $S = (\alpha, \beta)$

See the appendix for a proof.

The implication of this is that, intuitively enough, as confirmation bias assumes more importance relative to the instrumental payoff, the incentive to opt for a biased signal grows, and if ρ is sufficiently large, some set of A-biased consumers will have an optimal signal with $\beta < 1$.

However, the relationship between priors and signal preferences is somewhat different than in the pure case. This is because the $(\mu - \frac{1}{2})$ weighting on the psychological payoff, irrelevant in the pure case, matters here - it impacts the balance between the instrumental and psychological payoffs. Figure 1.6 graphs $\hat{\beta}(\mu; \rho)$ for various fixed values of ρ , showing how the set of consumers who prefer a biased signal grows as ρ increases. As the graph suggests, $\lim_{\mu \rightarrow \frac{1}{2}} \hat{\beta}(\mu; \rho) = 2$ for any ρ . This implies that for any arbitrarily large ρ , a sufficiently ambivalent consumer with μ close enough to $\frac{1}{2}$ will prefer a perfectly accurate signal. So will a sufficiently confident consumer with μ close enough to 1 or 0.

With the instrumental value of information present, the counter-intuitive result of the pure case, in which nearly-ambivalent consumers were the most biased, disappears. It is consumers ‘in the middle’ who may prefer a biased signal, with prior beliefs strong enough that they are psychologically invested in them and have significant confirmation bias, but not so strong that they would trust in a perfectly accurate information source to give them the news they want to hear.

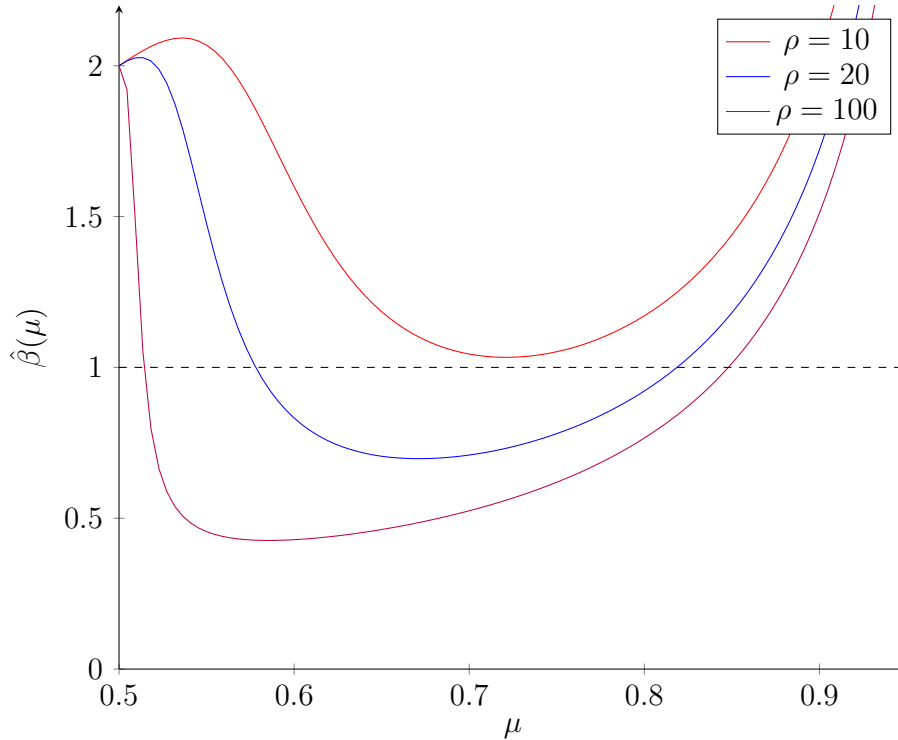


Figure 1.6: $\hat{\beta}(\mu; \rho)$ for various ρ

1.4.2 Firms

The situation faced by firms is qualitatively the same as in the case of pure confirmation bias. The translation from the distribution $F(\mu)$ of priors to the distribution $G(l)$ of consumers on the interval L is more complex, and consumers with different priors can have the same optimal signal (as is clear from Figure 1.6) and therefore be located at the same point on L . Nonetheless, the different distribution of consumers is essentially the only difference as regards the firms' problem, as the following results demonstrate.

Proposition 8. *With partially-biased preferences, if there exist two signals, $S_x, S_y \in L_s$ and a consumer located at a point l^* , such that either $l^* \geq l(S_x) > l(S_y)$ or $l^* \leq l(S_x) < l(S_y)$, that consumer prefers S_x to S_y .*

Proof. Assume a B-biased consumer, without loss of generality. For signals with the same direction of bias as l^* , of the form $(\alpha, 1)$, the result follows from the optimality of l^* and the concavity of $E(U|(\alpha, 1))$. For signals on the other half of L , of the form $(1, \beta)$, note that wherever $V_{pure}(S; \mu)$ is increasing in α or β so is $E(U|(\alpha, \beta))$, since $E(i_{match}|\alpha, \beta)$ is weakly increasing in both variables. The proof of Proposition 3 in the appendix demonstrates that $V_{pure}(S; \mu)$ is increasing in β (decreasing in distance from the perfectly-accurate centre point) along the other half of L , so the same is true of $E(U|(\alpha, \beta))$. \square

Proposition 9. *With partially-biased preferences, applying the 'trembling-hand-perfect' equilibrium refinement of Selten (1974) removes any equilibria in which firms offer 'off-the-line' signals with positive probability.*

Proof. It is sufficient to show that Lemma 5 holds under partially-biased preferences. The proof of Lemma 5 shows that for any signal $S_x = (\alpha_x, \beta_x)$, $S_x \notin L_s$, there is a signal $(\alpha_0, 1)$ such that $\alpha_0 > \alpha_x$ and $V_{pure}((\alpha_0, 1); \mu) > V_{pure}(S_x; \mu)$ for all $\mu \neq \frac{1}{2}$. Since $\alpha_0 > \alpha_x$, $1 \geq \beta_x$, and $E(i_{match}|\alpha, \beta)$ is weakly increasing in α and β , $E(i_{match}|\alpha_0, 1) \geq E(i_{match}|\alpha_x, \beta_x)$. Therefore $(\alpha_0, 1)$ is still universally preferred to (α_x, β_x) given partially-biased preferences and Lemma 5 still applies. \square

These results demonstrate that adding the instrumental value of information does not fundamentally alter the nature of the firms' problem, which can still be characterised as a one-dimensional location game. In fact, the proofs of Propositions 4, 5 and 6 apply without alteration. However, it should be stressed that while the game that results from the distribution of consumer preferences has the same structure in both the pure bias and partial bias cases, the results may be different, because the preference distribution itself will be different. In the pure case, a demand for more heavily biased signals would be driven by an initially uncertain population, but in the more plausible case of partial bias, consumers with somewhat stronger prior convictions would be the source of any demand for bias, as long as their beliefs are not so strong that they do not fear disconfirmation.

1.5 Conclusion

The model developed in this paper demonstrates how confirmation bias, expressed as preferences over future beliefs, admits a plausible formulation that leads to preferences over information structures consistent with the phenomenon of selective exposure. Unless they start out with sufficiently confident prior beliefs, consumers with preferences purely determined by confirmation bias prefer a biased information source to a perfectly accurate one. The bias can only ever be partial, since the source must sometimes report bad news in order to give the good news credibility, and the impact of good news on beliefs is weakened by the consumers' awareness of the bias. Nonetheless, this is a tradeoff a consumer may be willing to accept when the fact of confirmation matters more than the degree of it. Even when a competing preference for accurate information is present, as in Section 1.4, some group of consumers prefer biased signals as long as the weight placed on confirmation bias is large enough.

The relative weighting of confirmation bias¹¹ versus a preference for accuracy should not be thought of as fixed across all contexts, but as potentially varying depending on the subject matter that the state space represents, consistent with the mixed empirical results on the existence of selective exposure over different kinds of information. If confirmation bias has a larger impact when the information concerns political issues than when it concerns, say, entertainment news, this can be expressed in the model by a lower weight on confirmation bias.

The industrial setting that results from my formulation of consumer preferences, with heterogeneous priors leading to heterogeneous preferences that can be expressed spatially on an interval, is similar to Mullainathan and Shleifer (2005). The differing results for duopoly (in their model, the firms offer extreme bias in order to differentiate) should not be overemphasised, since their model includes prices where mine does not; the inclusion of prices leading to differentiation when firms would not differentiate without prices is a classic feature of Hotelling-style location models. The distinction between their model and the model of this paper lies in my account of consumers' beliefs and the relationship between those beliefs and their preferences, which provides clarity on just what kind of biased information they prefer, and how we should expect it to impact their beliefs. This could provide a basis on which to build an account of how beliefs and biased reporting develop together over time; one potential extension of this model is to introduce dynamics, and to consider under what circumstances differences in initial beliefs across consumers might be expected to converge or not.

In reality, there are a number of plausible explanations for the variation in reporting described in the introduction to this paper, and it is likely that a range of factors have some degree of influence. The results in this paper demonstrate that it is possible to give an account of confirmation bias as a potential cause of differing consumer tastes for reporting 'slants' that results in a familiar kind of product-differentiation problem for firms, without needing to admit any irrationality in how consumers process information and update their beliefs.

¹¹Represented in the model by the parameter ρ .

Bibliography

- [1] Bell, Jordan. (2014) “What is the domain of a solution of an ODE?”
<http://individual.utoronto.ca/jordanbell/>
- [2] Gentzkow, Matthew, and Jesse M. Shapiro. (2006) “Media bias and reputation.”
Journal of Political Economy 114.2: 280-316.
- [3] Groseclose, Tim, and Jeffrey Milyo. (2005) “A measure of media bias.” *The Quarterly Journal of Economics* 120.4: 1191-1237.
- [4] Hotelling, Harold. (1929) “Stability in competition.” *The Economic Journal* 39.153: 41-57.
- [5] Kamenica, Emir, and Matthew Gentzkow. (2011) “Bayesian persuasion.” *The American Economic Review* 101, no. 6: 2590-2615.
- [6] Knobloch-Westerwick, Silvia, and Jingbo Meng. (2009) “Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information.”
Communication Research 36.3: 426-448.
- [7] Lipnowski, Elliot, and Laurent Mathevet. (2018) “Disclosure to a Psychological Audience.” *American Economic Journal: Microeconomics* 10.4: 67-93.
- [8] Mullainathan, Sendhil, and Andrei Shleifer. (2005) “The Market for News.” *The American Economic Review* 95.4: 1031-1053.
- [9] Nickerson, Raymond S. (1998) “Confirmation bias: A ubiquitous phenomenon in many guises.” *Review of General Psychology* 2.2: 175.
- [10] Osborne, Martin J., and Carolyn Pitchik. (1986) “The nature of equilibrium in a location model.” *International Economic Review*: 223-237.
- [11] Pew Research Center. (2017) “Americans’ Attitudes About the News Media Deeply Divided Along Partisan Lines.”
- [12] Rabin, Matthew, and Joel L. Schrag. (1999) “First impressions matter: A model of confirmatory bias.” *The Quarterly Journal of Economics* 114.1: 37-82.

- [13] Smith, Matthew. (2017) “How left or right-wing are the UK’s newspapers?”
Yougov.co.uk
- [14] Stroud, Natalie Jomini. (2008) “Media use and political predispositions: Revisiting the concept of selective exposure.” *Political Behavior* 30.3: 341-366.

Appendix

Proof of Lemma 4

For $\mu > \frac{1}{2}$:

$$V_{pure}((1, \beta); \mu) = (\beta - (1 - \mu)\beta^2)^{\frac{1}{2}} - \beta(1 - \mu)^{\frac{1}{2}}$$

$$\frac{dV_{pure}((1, \beta); \mu)}{d\beta} = \frac{1 - 2(1 - \mu)\beta}{2(\beta - (1 - \mu)\beta^2)^{\frac{1}{2}}} - (1 - \mu)^{\frac{1}{2}}$$

$$\frac{d^2V_{pure}((1, \beta); \mu)}{d\beta^2} = \frac{-(1 - \mu)(\beta - (1 - \mu)\beta^2)^{\frac{1}{2}} - \frac{(1 - 2(1 - \mu)\beta)^2}{4(\beta - (1 - \mu)\beta^2)^{\frac{3}{2}}}}{(\beta - (1 - \mu)\beta^2)} < 0$$

Proof of Proposition 1

Lemma 3 implies that the consumer-optimal signal can be found by maximising the following function with respect to β :

$$V_{pure}((1, \beta)) = (\beta - \beta^2(1 - \mu))^{\frac{1}{2}} - \beta(1 - \mu)^{\frac{1}{2}}$$

Since this function is concave by Lemma 4, it has a unique maximum either at the boundary of its domain (which is $[0, 1]$ since β is a probability) or when the following first-order condition is satisfied.

$$\frac{1 - 2\beta(1 - \mu)}{2(\beta - \beta^2(1 - \mu))^{\frac{1}{2}}} = (1 - \mu)^{\frac{1}{2}}$$

That this condition is satisfied by $\frac{2 - \sqrt{2}}{4(1 - \mu)}$ is straightforwardly verified by substitution. The expression $\frac{2 - \sqrt{2}}{4(1 - \mu)}$ is increasing in μ , and greater than 1 if $\mu > \frac{2 + \sqrt{2}}{4}$, in which case the consumer-optimal signal is the perfectly-accurate signal with $\beta = 1$ by the concavity of $V_{pure}((1, \beta))$.

Proof of Proposition 3

Consider a B-biased consumer with l^* located on the left half of the interval. By the optimality of l^* and the concavity of $V_{pure}((\alpha, 1); \mu)$, moving in either direction along the interval (increasing or decreasing α) decreases the value of the signal, up to the left endpoint ($l = \frac{1}{2} - \frac{1}{2\sqrt{2}}$) or to the perfectly-accurate centre point ($l = \frac{1}{2}$). This establishes the result for

$l_y < l^*$. For $l_y > l^*$ we must also consider the other half of the interval - signals of the form $(1, \beta)$. Note that for $\mu < \frac{1}{2}$:

$$\begin{aligned} V_{pure}((1, \beta); \mu) &= \beta^{\frac{1}{2}} \left[((1 - \mu)\beta)^{\frac{1}{2}} - (1 - (1 - \mu)\beta)^{\frac{1}{2}} \right] \\ &= - \left[(\beta - (1 - \mu)\beta^2)^{\frac{1}{2}} - \beta(1 - \mu)^{\frac{1}{2}} \right] \end{aligned}$$

By Lemmas 4 and 1, this is the negative of a concave function that is maximised at $\frac{2-\sqrt{2}}{4(1-\mu)}$. Therefore, it is a convex function that is minimised at $\frac{2-\sqrt{2}}{4(1-\mu)}$. For $\mu < \frac{1}{2}$, $\frac{2-\sqrt{2}}{4(1-\mu)} > 1 - \frac{1}{\sqrt{2}}$; that is, the value of β that minimises $V_{pure}((1, \beta); \mu)$ is lower than the value of β corresponding to any point on the right half of the interval (the minimum is 'beyond' the endpoint of the interval), so $V_{pure}((1, \beta); \mu)$ is increasing in β at any such point. This means that the value of l_y decreases as it moves rightwards along the interval away from the perfectly accurate centre point. This establishes the result for $l_y > l^*$.

The proof for A-biased consumers is analogous.

Proof of Lemma 5

Consider a signal $S_x = (\alpha_x, \beta_x)$, $S_x \notin L_s$. Assume without loss of generality that $\alpha_x \leq \beta_x$ (noting that in combination with the labelling restriction $\beta > 1 - \alpha$, this implies that $\beta \geq \frac{1}{2}$).

I will be concerned with the following function:

$$\begin{aligned} V_0(\alpha, \beta) &= \lim_{\mu \downarrow \frac{1}{2}} V_{pure}((\alpha, \beta); \mu) \\ &= - \lim_{\mu \uparrow \frac{1}{2}} V_{pure}((\alpha, \beta); \mu) \\ &= \frac{1}{\sqrt{2}} (\alpha + \beta - 1)^{\frac{1}{2}} ((\alpha + 1 - \beta)^{\frac{1}{2}} - (\beta + 1 - \alpha)^{\frac{1}{2}}) \\ &= \frac{1}{\sqrt{2}} ((\alpha^2 - (1 - \beta)^2)^{\frac{1}{2}} - (\beta^2 - (1 - \alpha)^2)^{\frac{1}{2}}) \end{aligned}$$

I will show that there exists a value α_0 with $\alpha_x < \alpha_0 \leq 1$, such that $V_0(\alpha, \beta) = V_0(\alpha_0, 1)$, and that the signal $(\alpha_0, 1)$ is preferred to (α_x, β_x) by all consumers.

Lemma A.1. *For $\alpha \leq \beta$, $\frac{\partial V_0}{\partial \beta} < 0$, except at $\alpha = \beta = \frac{1}{2}$.*

Proof. For $\alpha < \beta$, $((\alpha + 1 - \beta)^{\frac{1}{2}} - (\beta + 1 - \alpha)^{\frac{1}{2}})$ is negative. It is also decreasing in β . Since $V_0(\alpha, \beta)$ is the product of a term that is positive and increasing in β and a term that is negative and decreasing in β , it is decreasing in β .

For $\alpha = \beta$:

$$\frac{\partial V_0}{\partial \beta} = \frac{1 - \beta}{(\beta^2 - (1 - \beta)^2)^{\frac{1}{2}}} - \frac{\beta}{(\beta^2 - (1 - \beta)^2)^{\frac{1}{2}}} < 0$$

This partial derivative is not defined at $\alpha = \beta = \frac{1}{2}$, but the limit as α and β approach $\frac{1}{2}$ is 0. □

Lemma A.2. *For any signal $S_x = (\alpha_x, \beta_x)$ with $\alpha_x \leq \beta_x$ that is not in L_s , there exists a unique $\alpha_0 > \alpha_x$ such that $V_0(\alpha_0, 1) = V_0(\alpha_x, \beta_x)$ and $\alpha_0 > 1 - \frac{1}{\sqrt{2}}$.*

Proof. Assume that $\beta_x < 1$. By Lemma A.1:

$$V_0(\alpha_x, 1) < V_0(\alpha_x, \beta_x)$$

Since $\alpha_x \leq \beta_x$, $V_0(\alpha_x, \beta_x) \leq 0$. The function $V_0(\alpha, 1) = \frac{1}{\sqrt{2}}(\alpha - (2\alpha - \alpha^2)^{\frac{1}{2}})$ is convex and continuous in α for $0 \leq \alpha \leq 2$, and $V_0(1, 1) = 0$. We have:

$$V_0(\alpha_x, 1) < V_0(\alpha_x, \beta_x) \leq V_0(1, 1)$$

By the intermediate value theorem, there exists a value α_0 in the interval $(\alpha_x, 1]$ such that $V_0(\alpha_0, 1) = V_0(\alpha_x, \beta_x)$. This value is unique by the convexity of $V_0(\alpha, 1)$.

Now assume that $\beta_x = 1$. Since $S_x \notin L_s$, this implies that $\alpha_x < 1 - \frac{1}{\sqrt{2}}$. The convex function $V_0(\alpha, 1)$ has its minimum at $\alpha = 1 - \frac{1}{\sqrt{2}}$. Since $V_0(\alpha_x, 1) \leq 0$ and $V_0(1, 1) = 0$, the intermediate value theorem again ensures that an appropriate α_0 exists. □

Lemma A.3. *For $\mu > \frac{1}{2}$, the cross partial derivatives $\frac{\partial^2 V_{pure}((\alpha, \beta); \mu)}{\partial \alpha \partial \mu}$ and $\frac{\partial^2 V_{pure}((\alpha, \beta); \mu)}{\partial \beta \partial \mu}$ are always positive. For $\mu < \frac{1}{2}$, they are always negative.*

Proof. I use the expressions $P(a)$ and $P(b)$ for the probabilities of the possible reports in the algebra of this proof, for readability. To follow their use, note the following equalities:

$$\begin{aligned} P(a) &= \mu(\alpha + \beta - 1) + 1 - \beta \\ P(b) &= 1 - P(a) = (1 - \mu)(\alpha + \beta - 1) + 1 - \alpha \\ \frac{\partial P(a)}{\partial \mu} &= \alpha + \beta - 1 \\ \frac{\partial P(a)}{\partial \alpha} &= \mu \\ \frac{\partial P(a)}{\partial \beta} &= -(1 - \mu) \end{aligned}$$

The main proof is as follows. For $\mu > \frac{1}{2}$:

$$\begin{aligned}
V_{pure}(\alpha, \beta; \mu) &= (\alpha + \beta - 1)^{\frac{1}{2}} (P(a)^{\frac{1}{2}} - P(b)^{\frac{1}{2}}) \\
\frac{\partial V_{pure}(\alpha, \beta; \mu)}{\partial \mu} &= \frac{1}{2}(\alpha + \beta - 1)^{\frac{3}{2}} \left(\frac{1}{P(a)^{\frac{1}{2}}} + \frac{1}{P(b)^{\frac{1}{2}}} \right) \\
\frac{\partial^2 V_{pure}(\alpha, \beta; \mu)}{\partial \mu \partial \alpha} &= \frac{1}{2} \left(\frac{\frac{3}{2}(\alpha + \beta - 1)^{\frac{1}{2}} P(a)^{\frac{1}{2}} - \frac{1}{2} \mu (\alpha + \beta - 1)^{\frac{3}{2}} P(a)^{-\frac{1}{2}}}{P(a)} \right. \\
&\quad \left. + \frac{\frac{3}{2}(\alpha + \beta - 1)^{\frac{1}{2}} P(b)^{\frac{1}{2}} + \frac{1}{2} \mu (\alpha + \beta - 1)^{\frac{3}{2}} P(b)^{-\frac{1}{2}}}{P(b)} \right) \\
&= \frac{1}{4}(\alpha + \beta - 1)^{\frac{1}{2}} \left(\frac{3P(a) - \mu(\alpha + \beta - 1)}{P(a)^{\frac{3}{2}}} + \frac{3P(b) + \mu(\alpha + \beta - 1)}{P(b)^{\frac{3}{2}}} \right) \\
&= \frac{1}{4}(\alpha + \beta - 1)^{\frac{1}{2}} \left(\frac{2P(a) + 1 - \beta}{P(a)^{\frac{3}{2}}} + \frac{3P(b) + \mu(\alpha + \beta - 1)}{P(b)^{\frac{3}{2}}} \right) > 0 \\
\frac{\partial^2 V_{pure}(\alpha, \beta; \mu)}{\partial \mu \partial \beta} &= \frac{1}{4}(\alpha + \beta - 1)^{\frac{1}{2}} \left(\frac{3P(a) + (1 - \mu)(\alpha + \beta - 1)}{P(a)^{\frac{3}{2}}} \right. \\
&\quad \left. + \frac{3P(b) - (1 - \mu)(\alpha + \beta - 1)}{P(b)^{\frac{3}{2}}} \right) \\
&= \frac{1}{4}(\alpha + \beta - 1)^{\frac{1}{2}} \left(\frac{3P(a) + (1 - \mu)(\alpha + \beta - 1)}{P(a)^{\frac{3}{2}}} + \frac{2P(b) + 1 - \alpha}{P(b)^{\frac{3}{2}}} \right) \\
&> 0
\end{aligned}$$

For $\mu < \frac{1}{2}$, note that:

$$\frac{\partial V_{pure}(\alpha, \beta; \mu)}{\partial \mu} = -\frac{1}{2}(\alpha + \beta - 1)^{\frac{3}{2}} \left(\frac{1}{P(a)^{\frac{1}{2}}} + \frac{1}{P(b)^{\frac{1}{2}}} \right)$$

Therefore the cross partial derivatives in this case are given by the negatives of the expressions derived above, and are negative.

□

Lemma A.4. *The signal $(\alpha_0, 1)$ is strictly preferred to (α_x, β_x) by every consumer with $\mu \neq \frac{1}{2}$.*

Proof. The difference in the value of the two signals is given by:

$$\begin{aligned}
V_{pure}((\alpha_0, 1); \mu) - V_{pure}((\alpha_x, \beta_x); \mu) &= V_{pure}((\alpha_0, 1); \mu) - V_{pure}((\alpha_x, 1); \mu) \\
&\quad + V_{pure}((\alpha_x, 1); \mu) - V_{pure}((\alpha_x, \beta_x); \mu)
\end{aligned}$$

Differentiating with respect to the prior gives:

$$\begin{aligned} \frac{\partial}{\partial \mu} (V_{pure}((\alpha_0, 1); \mu) - V_{pure}((\alpha_x, \beta_x); \mu)) &= \frac{\partial V}{\partial \mu}((\alpha_0, 1); \mu) - \frac{\partial V}{\partial \mu}((\alpha_x, 1); \mu) \\ &\quad + \frac{\partial V}{\partial \mu}((\alpha_x, 1); \mu) - \frac{\partial V}{\partial \mu}((\alpha_x, \beta_x); \mu) \end{aligned}$$

Recall that $\alpha_0 > \alpha_x$. Since $\beta_x \leq 1$, applying Lemma A.3 gives, for $\mu > \frac{1}{2}$, $\frac{\partial V}{\partial \mu}((\alpha_0, 1); \mu) > \frac{\partial V}{\partial \mu}((\alpha_x, 1); \mu)$ and $\frac{\partial V}{\partial \mu}((\alpha_x, 1); \mu) \geq \frac{\partial V}{\partial \mu}((\alpha_x, \beta_x); \mu)$; so the derivative above is positive. For $\mu < \frac{1}{2}$, the inequalities are reversed, and the derivative is negative.

We have, therefore, that $V_{pure}((\alpha_0, 1); \mu) - V_{pure}((\alpha_x, \beta_x); \mu)$ is increasing in μ for $\mu > \frac{1}{2}$ and decreasing in μ for $\mu < \frac{1}{2}$. We also have:

$$\lim_{\mu \downarrow \frac{1}{2}} \left(V_{pure}((\alpha_0, 1); \mu) - V \right) = V_0((\alpha_0, 1); \mu) - V_0((\alpha_x, \beta_x)) = 0$$

$$\lim_{\mu \uparrow \frac{1}{2}} \left(V_{pure}((\alpha_0, 1); \mu) - V_{pure}((\alpha_x, \beta_x); \mu) \right) = - \left(V_0((\alpha_0, 1); \mu) - V_0((\alpha_x, \beta_x); \mu) \right) = 0$$

Therefore, $V_{pure}((\alpha_0, 1); \mu) > V_{pure}((\alpha_x, \beta_x))$ for any $\mu \neq \frac{1}{2}$.

□

Proof of Corollary 2

I first prove that β_{match}^* is decreasing in ρ .

$$\begin{aligned} -\frac{d\beta_{match}^*}{d\rho} &= \frac{d}{d\rho} \left(\frac{\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) - 1}{(2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{\frac{1}{2}}} \right) \\ &= \frac{\mu^{\frac{1}{2}}(\mu - \frac{1}{2})(2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{\frac{1}{2}}}{2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1} \\ &\quad - \frac{(2\rho\mu(\mu - \frac{1}{2})^2 - \mu^{\frac{1}{2}}(\mu - \frac{1}{2}))(2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{-\frac{1}{2}}(\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) - 1)}{2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1} \\ &= \frac{\rho\mu(\mu - \frac{1}{2})^2}{(2\rho^2\mu(\mu - \frac{1}{2})^2 - 2\rho\mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{\frac{3}{2}}} > 0 \\ \frac{d\beta_{match}^*}{d\rho} &< 0 \end{aligned}$$

To prove that β_{match}^* approaches β_{bias}^* as ρ goes to infinity, I show that β_{match}^* approaches $\frac{1}{2(1-\mu)}(1 - \frac{1}{\sqrt{2}})$. Taking the limit of the relevant expression in ρ :

$$\begin{aligned}
& \lim_{\rho \rightarrow \infty} \frac{\rho \mu^{\frac{1}{2}}(\mu - \frac{1}{2}) - 1}{(2\rho^2 \mu(\mu - \frac{1}{2})^2 - 2\rho \mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{\frac{1}{2}}} \\
&= \lim_{\rho \rightarrow \infty} \left(\frac{\mu^{\frac{1}{2}}(\mu - \frac{1}{2})}{(2\mu(\mu - \frac{1}{2})^2 - 2\rho^{-1} \mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + \rho^{-2})^{\frac{1}{2}}} - \frac{1}{(2\rho^2 \mu(\mu - \frac{1}{2})^2 - 2\rho \mu^{\frac{1}{2}}(\mu - \frac{1}{2}) + 1)^{\frac{1}{2}}} \right) \\
&= \frac{\mu^{\frac{1}{2}}(\mu - \frac{1}{2})}{(2\mu(\mu - \frac{1}{2})^2)^{\frac{1}{2}}} \\
&= \frac{1}{\sqrt{2}}
\end{aligned}$$

Taking the limit of β_{match}^* and applying the above result gives:

$$\lim_{\rho \rightarrow \infty} \beta_{match}^* = \frac{1}{2(1-\mu)}(1 - \frac{1}{\sqrt{2}})$$

Proof of Proposition 6

Let $S(l)$ be the unique signal located at l and let $V(l, \mu)$ be an expression giving the value of that signal to a consumer with prior $\mu \in (0, 1)$. Specifically, with purely confirmation biased preferences, as in Section 1.3:

$$V(l, \mu) = V_{pure}(S(l); \mu)$$

In the case of state-matching preferences, as in Section 1.4:

$$V(l, \mu) = E(U|S, \mu)$$

In order to specify a firm's expected payoff when locating at a given point, I introduce a function, $C(l, l') : L^2 \rightarrow [0, 1]$, which when $l \neq l'$ gives the probability that a consumer prefers the left-most of the two signals, l and l' . At $l = l'$ it gives the probability that $l^*(\mu) \leq l$; that is, $G(l)$.

The methods I use to prove Proposition 6 require that for a fixed $\bar{l} \neq 0.5$, $C(\bar{l}, l')$ is continuous and differentiable in l' except at a finite number of points. The various lemmas that follow contribute to a proof of this, culminating in Lemmas A.19 and A.20. Note that I use] [as

the notation for open endpoints of an interval in this section, to avoid confusion with ordered pairs.

I begin by defining $C(l, l')$:

$$C(l, l') = \begin{cases} P(V(l', \mu) \geq V(l, \mu)) & l' < l \\ G(l) & l' = l \\ P(V(l, \mu) \geq V(l', \mu)) & l' > l \end{cases}$$

Lemma A.5. *For any $\bar{l} \in L \setminus \{0.5\}$, $C(\bar{l}, l')$ is continuous at $l' = \bar{l}$.*

Proof. I first show that $\lim_{l' \downarrow \bar{l}} C(\bar{l}, l') = G(\bar{l})$:

$$\begin{aligned} \lim_{l' \downarrow \bar{l}} C(\bar{l}, l') &= \lim_{l' \downarrow \bar{l}} P(V(\bar{l}, \mu) \geq V(l', \mu)) \\ &= \lim_{l' \downarrow \bar{l}} [P(V(\bar{l}, \mu) \geq V(l', \mu) | l^*(\mu) \leq \bar{l}) P(l^*(\mu) \leq \bar{l}) \\ &\quad + P(V(\bar{l}, \mu) \geq V(l', \mu) | \bar{l} < l^*(\mu) \leq l') P(\bar{l} < l^*(\mu) \leq l')] \\ &= \lim_{l' \downarrow \bar{l}} [P(l^*(\mu) \leq \bar{l}) + P(V(\bar{l}, \mu) \geq V(l', \mu) | \bar{l} < l^*(\mu) \leq l') P(\bar{l} < l^*(\mu) \leq l')] \\ &= P(l^*(\mu) \leq \bar{l}) \\ &= G(\bar{l}) \end{aligned}$$

Here the second and third equalities use the facts that for $l' > \bar{l}$:

$$\begin{aligned} P(V(\bar{l}, \mu) \geq V(l', \mu) | l^*(\mu) > l') &= 0 \\ P(V(\bar{l}, \mu) \geq V(l', \mu) | l^*(\mu) \leq \bar{l}) &= 1 \end{aligned}$$

These follow from Propositions 3 and 8.

A similar argument shows that $\lim_{l' \uparrow \bar{l}} C(\bar{l}, l') = G(\bar{l})$:

$$\begin{aligned} \lim_{l' \uparrow \bar{l}} C(\bar{l}, l') &= \lim_{l' \uparrow \bar{l}} P(V(l', \mu) \geq V(\bar{l}, \mu)) \\ &= \lim_{l' \uparrow \bar{l}} [P(V(l', \mu) \geq V(\bar{l}, \mu) | l^*(\mu) \leq \bar{l}) P(l^*(\mu) \leq l') \\ &\quad + P(V(l', \mu) \geq V(\bar{l}, \mu) | \bar{l} < l^*(\mu) \leq l') P(\bar{l} < l^*(\mu) \leq l')] \\ &= \lim_{l' \uparrow \bar{l}} [P(l^*(\mu) \leq l') + P(V(l', \mu) \geq V(\bar{l}, \mu) | l' < l^*(\mu) \leq \bar{l}) P(l' < l^*(\mu) \leq \bar{l})] \\ &= \lim_{l' \uparrow \bar{l}} P(l^*(\mu) \leq l') \\ &= \lim_{l' \uparrow \bar{l}} G(l') \\ &= G(\bar{l}) \end{aligned}$$

Here the last equality uses the continuity of $G(l)$ for $l \neq 0.5$. □

I now specify $C(l, l')$ in terms of the values of μ at which the sign of $V(l, \mu) > V(l', \mu)$ changes. Let $Z(l, l') = (Z_1(l, l'), Z_2(l, l') \dots Z_n(l, l'))$ be the ascending-ordered ($Z_1(l, l') > Z_2(l, l') > \dots > Z_n(l, l')$) set of such values. Note that for $l \neq l'$ this set is finite (and non-empty, since $l, l' \in L$): since $V(l, \mu) - V(l', \mu)$ is continuous in μ except at 0.5, $Z \setminus \{0.5\}$ is a subset of the zeroes of that function, which are themselves a subset of the zeroes of a polynomial in μ .

The specification of $C(l, l')$ in terms of the members of $Z(l, l')$ will depend on whether the sign of $V(l, \mu) > V(l', \mu)$ begins and ends positive or negative at the boundaries of $[0, 1]$. For ease of notation I define a pair of indicator functions:

$$I_0(l, l') = \begin{cases} 0 & V(l, \mu) - V(l', \mu) \leq 0, \forall \mu \in [0, Z_1(l, l')[\\ 1 & \text{otherwise} \end{cases}$$

$$I_1(l, l') = \begin{cases} 0 & V(l, \mu) - V(l', \mu) \leq 0, \forall \mu \in]Z_n(l, l'), 1] \\ 1 & \text{otherwise} \end{cases}$$

I now specify $C(l, l')$ in terms of the members of $Z(l, l')$ as follows:

$$C(l, l') = \begin{cases} I_1(l', l) + \sum_{Z_i \in Z(l', l)} (-1)^{i-I_0(l', l)} F(Z_i) & l' < l \\ G(l) & l' = l \\ I_1(l, l') + \sum_{Z_i \in Z(l, l')} (-1)^{i-I_0(l, l')} F(Z_i) & l' > l \end{cases}$$

As noted above, the members of $Z(\bar{l}, l')$ (other than 0.5 if it is a member of the set) are also zeroes of $V(\bar{l}, \mu) - V(l', \mu)$. In fact, all zeroes of this function at a given l' are members of Z unless they are also local extrema of the function, at which $\frac{d}{d\mu} V(\bar{l}, \mu) = \frac{d}{d\mu} V(l', \mu)$. Using this fact, it is possible to show that on certain domains, the non-0.5 members of $Z(\bar{l}, l')$ are given by functions that are differentiable in l' ; specifically, by the solutions to particular initial value problems.

I begin by defining the appropriate domain for these functions:

$$X_1 = \left\{ (l', \mu) \in ((\bar{l}, 1) \times ([0, 0.5[\cup]0.5, 1])) : \nexists \frac{d}{d\mu} V(l', \mu) \right\}$$

$$X_2 = \left\{ (l', \mu) \in ((\bar{l}, 1) \times ([0, 0.5[\cup]0.5, 1])) : \frac{d}{d\mu} V(\bar{l}, \mu) - \frac{d}{d\mu} V(l', \mu) = 0 \right\}$$

$$E = ((\bar{l}, 1) \times ([0, 0.5[\cup]0.5, 1]) \setminus (X_1 \cap X_2))$$

Here X_1 is the set of points (l', μ) at which V is not differentiable in l' . Note that all points at which $l' = 0.5$ are in X_1 , and when preferences are given by the state-matching utility function X_1 also contains points at which a posterior belief of 0.5 is possible¹². X_2 is the set of points (l', μ) at which the function $V(\bar{l}, \mu) - V(l', \mu)$ has a local extremum in μ . E is the set of points (l', μ) which are not in X_1 or X_2 and do not have l' equal to 0, 0.5, or 1.

Define the function $g(l', \mu) : E \rightarrow \mathbb{R}$ as follows:

$$g(l', \mu) = \frac{\frac{d}{dl} V(l', \mu)}{\frac{d}{d\mu} V(\bar{l}, \mu) - \frac{d}{d\mu} V(l', \mu)}$$

At any given $l' = l'_0$, I define a function $z_i(l'; \bar{l}, l'_0)$ corresponding to each $Z_i(\bar{l}, l'_0) \in Z(\bar{l}, l'_0)$ as the maximal solution (i.e. having the maximal interval of validity) to the ordinary differential equation:

$$\frac{d}{dl'} z_i(l'; \bar{l}, l'_0) = g(l', z_i(l'; \bar{l}, l'_0))$$

with initial condition $z_i(l'_0; \bar{l}, l'_0) = Z_i(l, l')$. The function exists and is unique by the Picard-Lindelöf theorem. Denote its domain by:

$$\omega(z_i(l'_0; \bar{l}, l'_0)) =]\omega_-(z_i), \omega_+(z_i)[$$

By the extension theorem for IVP solutions¹³, as each function $z_i(l'; \bar{l}, l'_0)$ approaches either boundary of its domain, $(l', z_i(l'; \bar{l}, l'_0))$ approaches the boundary of E .

I define the following sets:

$$\hat{Z}(\bar{l}, l'_0) = \left\{ z_i(l'; \bar{l}, l'_0) : Z_i(\bar{l}, l'_0) = z_i(l'_0; \bar{l}, l'_0) \right\}$$

$$\bar{Z} = \left\{ z_i(l'; \bar{l}, l') : \exists l' \in (\bar{l}, 1), Z_i(\bar{l}, l') = z_i(l'; \bar{l}, l') \right\}$$

$\hat{Z}(\bar{l}, l'_0)$ contains all the z functions generated by the members of $Z(\bar{l}, l'_0)$ at a given l'_0 (which is every z function for which $l'_0 \in \omega(z)$), and \bar{Z} contains all such functions generated at any point in $(\bar{l}, 1)$.

Lemma A.6. *Two functions $z_i(l'; \bar{l}, l'_0)$ and $z_j(l'; \bar{l}, l'_0)$ with $i \neq j$ do not intersect at any point.*

Proof. First note that the two functions are distinct at l'_0 :

$$z_i(l'_0; \bar{l}, l'_0) \neq z_j(l'_0; \bar{l}, l'_0)$$

¹²Because at such points the signal switches between affecting the consumer's action or not, causing a jump in $\frac{dV}{dl}$.

¹³See Bell (2014) for a discussion of these theorems and further references.

Now assume that the two functions intersect at some point x :

$$z_i(x; \bar{l}, l'_0) = z_j(x; \bar{l}, l'_0) = Z_h(\bar{l}, x)$$

Consider a z function defined at x , $z_h(l'; \bar{l}, x)$. On any interval containing x and within the domain of $z_h(l'; \bar{l}, x)$, the restriction of $z_h(l'; \bar{l}, x)$ to that interval is the unique solution to the initial value problem that defines $z_h(l'; \bar{l}, x)$. Restrictions of the original two functions also solve the initial value problem on any such interval that falls within their domain, so the three functions are equal at any point that falls within all three of their domains; the original two functions must diverge outside of the domain of $z_h(l'; \bar{l}, x)$.

However, since the restrictions of the original two functions to any interval that falls within their domains and includes x both solve the initial value problem at x , then if those functions are defined beyond the bounds of $z_h(l'; \bar{l}, x)$ then the latter function does not have the maximal interval of validity for solutions to the initial value problem at x , which contradicts its definition. Therefore the point of intersection x cannot exist. \square

Call an open interval $]a, b[\in [\bar{l}, 1]$ a **fixed-Z interval** if both of the following hold for any $l' \in]a, b[$:

$$\left\{ \bar{\mu} \in [0, 1] : \lim_{\mu \uparrow \bar{\mu}} V(\bar{l}, \mu) - V(l', \mu) = 0 \cap (l', \bar{\mu}) \notin E \right\} = \emptyset$$

$$\left\{ \bar{\mu} \in [0, 1] : \lim_{\mu \downarrow \bar{\mu}} V(\bar{l}, \mu) - V(l', \mu) = 0 \cap (l', \bar{\mu}) \notin E \right\} = \emptyset$$

That is, a point l' is only contained in a fixed-Z interval if any pair (l', μ) that is a zero of $V(\bar{l}, \mu) - V(l', \mu)$ is in E , both the left and right limits at 0.5 of $V(\bar{l}, \mu) - V(l', \mu)$ are non-zero, and the limits of the same expression at 0 and 1 are also non-zero. Note that this definition implies that on the interior of a fixed-Z interval, all elements of Z not equal to 0.5 are given by z functions, since each non-0.5 element of $Z(\bar{l}, l')$ must be a zero of $V(\bar{l}, \mu) - V(l', \mu)$.

The next two Lemmas demonstrate that the elements of Z and their ordering are consistent over such intervals, in the sense that if a particular element Z_i is given by a particular z function anywhere on a fixed-Z interval, it is given by the same function throughout the interval, and if 0.5 is the j th element of Z anywhere on such an interval, it is the j th element of Z over the entire interval.

I begin by establishing that if a z function is defined anywhere on a fixed-Z interval, it is defined on the whole interval.

Lemma A.7. *For any function $z \in \bar{Z}$, neither boundary of its domain is contained in a fixed-Z interval.*

Proof. Consider the upper boundary of the domain of z , $\omega_+(z)$. Let z_+ denote the limit of z at that boundary, i.e:

$$z_+ = \lim_{l' \uparrow \omega_+(z)} z$$

By the extension theorem:

$$(\omega_+(z), z_+) \in \partial E$$

Since E is an open set, this implies that:

$$(\omega_+(z), z_+) \notin E$$

Since at each l' in the domain of z , $V(\bar{l}, z(l')) - V(l', z(l')) = 0$, it is clear that:

$$\lim_{\mu \uparrow z_+} V(\bar{l}, \mu) - V(\omega_+(z), \mu) = 0$$

It follows from the definition of a fixed-Z interval that $\omega_+(z)$ is not contained in one. The argument for the lower boundary $\omega_-(z)$ is analogous. \square

Lemmas A.6 and A.7 together imply that the same set of z functions are defined at any point on a fixed-Z interval, and that their ordering is also consistent over the interval; i.e. if $]a, b[$ is a fixed-Z interval and $x, x' \in]a, b[$, then $z_i(x; \bar{l}, x) > z_j(x; \bar{l}, x)$ if and only if $z_i(x'; \bar{l}, x) > z_j(x'; \bar{l}, x)$.

Lemma A.8. *Let $]a, b[$ be a fixed-Z interval. Then either $0.5 \in Z(\bar{l}, l')$ for all $l' \in]a, b[$, or $0.5 \notin Z(\bar{l}, l')$ for all $l' \in]a, b[$.*

Proof. Assume that two points $x, x' \in]a, b[$ exist such that $0.5 \in Z(\bar{l}, x)$ and $0.5 \notin Z(\bar{l}, x')$, and that x can be chosen such that $0.5 = Z_i(\bar{l}, x)$, $i \neq 1, i \neq n$, where $n = |Z(\bar{l}, x)|$. This implies that functions $z_{i-1}(l'; \bar{l}, x)$ and $z_{i+1}(l'; \bar{l}, x)$ exist and that $\frac{d}{d\mu} V(\bar{l}, z_{i-1}(x; \bar{l}, x)) - \frac{d}{d\mu} V(x, z_{i-1}(x; \bar{l}, x))$ and $\frac{d}{d\mu} V(\bar{l}, z_{i+1}(x; \bar{l}, x)) - \frac{d}{d\mu} V(x', z_{i+1}(x; \bar{l}, x))$ have the same sign.

Since $0.5 \notin Z_i(\bar{l}, x')$, $\frac{d}{d\mu} V(\bar{l}, z_{i-1}(x'; \bar{l}, x)) - \frac{d}{d\mu} V(x', z_{i-1}(x'; \bar{l}, x))$ and $\frac{d}{d\mu} V(\bar{l}, z_{i+1}(x'; \bar{l}, x)) - \frac{d}{d\mu} V(x', z_{i+1}(x'; \bar{l}, x))$ must have opposite sign. This implies that one of these two differences in derivatives must change sign on the interval $[x, x']$. The z functions do not go to 0.5 on a fixed-Z interval (any point $(l', 0.5)$ is in the boundary of E), so the derivatives are continuous. However, the difference in the derivatives also does not go to 0 on a fixed-Z interval. Therefore they cannot change sign.

Now assume that x and x' exist as before, but x cannot be chosen such that $0.5 = Z_i(\bar{l}, x)$, $i \neq 1, i \neq n$. This implies that either $0.5 = Z_1(\bar{l}, x)$ or $0.5 = Z_n(\bar{l}, x)$. It cannot be both, which would make it the only element of $Z(\bar{l}, x)$: since $0.5 \notin Z(\bar{l}, x')$ and $Z(\bar{l}, x')$ is non-empty, there exists a function $z_1(l'; \bar{l}, x')$, and by Lemma A.7, $z_1(x; \bar{l}, x') \in Z(\bar{l}, x)$. Assume without loss of generality that $x < x'$, that $0.5 = Z_1(\bar{l}, l') \forall l' \in [x, x']$, and that $\frac{d}{d\mu} V(\bar{l}, z_1(x'; \bar{l}, x')) - \frac{d}{d\mu} V(x', z_1(x'; \bar{l}, x')) < 0$. Since $\frac{d}{d\mu} V(\bar{l}, z_1(l'; \bar{l}, x')) - \frac{d}{d\mu} V(l', z_1(l'; \bar{l}, x'))$ cannot change sign on a fixed-Z interval, these assumptions imply that for $l' \in [x, x']$, $V(\bar{l}, \mu) - V(l', \mu) \leq 0 \forall \mu \in (0, 0.5)$ and that $V(\bar{l}, \mu) - V(x', \mu) \geq 0, \forall \mu \in (0, z_1(x'; \bar{l}, x'))$. From this and the continuity of $V(\bar{l}, \mu) - V(l', \mu)$ in l' for $\mu \neq 0.5$ it follows that $V(\bar{l}, \mu) - V(x', \mu) = 0, \forall \mu \in (0, 0.5)$. This clearly cannot hold for $x' \neq \bar{l}$, for either specification of $V(l, \mu)$. \square

The preceding lemmas show that over a fixed-Z interval, any particular Z_i is consistently equal to either a particular z function or to 0.5. I now show that the remaining elements of the specification of $C(\bar{l}, l')$ are constant over a fixed-Z interval.

Lemma A.9. *The values of $I_0(\bar{l}, l')$ and $I_1(\bar{l}, l')$ are constant on any fixed-Z interval.*

Proof. Assume that a fixed-Z interval $]a, b[$ exists such that for two points $x, x' \in]a, b[, x < x'$, $I_0(\bar{l}, x) \neq I_0(\bar{l}, x')$. Assume further that $Z_1(\bar{l}, x) \neq 0.5$, so that the function $z_1(x'; \bar{l}, x')$ exists. Since $]a, b[$ is a fixed-Z interval, $z_1(x'; \bar{l}, x') = z_1(x'; \bar{l}, x) \neq 0.5$. Since $I_0(\bar{l}, x) \neq I_0(\bar{l}, l')$, this implies that $\frac{d}{d\mu} V(\bar{l}, z_1(x; \bar{l}, x)) - \frac{d}{d\mu} V(x, z_1(x; \bar{l}, x))$ and $\frac{d}{d\mu} V(\bar{l}, z_1(x'; \bar{l}, x)) - \frac{d}{d\mu} V(x', z_1(x'; \bar{l}, x))$ must have opposite sign. Since $\frac{d}{d\mu} V(\bar{l}, \mu) - \frac{d}{d\mu} V(l', \mu)$ is continuous in μ except at 0.5 and $z_1(l'; \bar{l}, l'_0)$ is continuous, this change of sign requires either that $\frac{d}{d\mu} V(\bar{l}, z_1(l'; \bar{l}, x)) - \frac{d}{d\mu} V(l', z_1(l'; \bar{l}, x))$ goes to 0 at some point on $[x, x']$ or that $z_1(l'; \bar{l}, x)$ goes to 0.5 at some point on $[x, x']$. Either contradicts the assumption that $]a, b[$ is a fixed-Z interval.

Now assume that $Z_1(\bar{l}, x) = 0.5$. Since $]a, b[$ is a fixed-Z interval, this implies that $Z_1(\bar{l}, l') = 0.5 \forall l' \in]a, b[$. Assume without loss of generality that $I_0(\bar{l}, l') = I_0(\bar{l}, x') = 1, \forall l' \in [x, x')$. This implies that for $l' \in [x, x'[$, $V(\bar{l}, \mu) - V(l', \mu) \geq 0, \forall \mu \in]0, 0.5[$ and that $V(\bar{l}, \mu) - V(x', \mu) \leq 0, \forall \mu \in]0, 0.5[$. From this and the continuity of $V(\bar{l}, \mu) - V(l', \mu)$ in l' for $\mu \neq 0.5$ it follows that $V(\bar{l}, \mu) - V(x', \mu) = 0, \forall \mu \in]0, 0.5[$. This cannot hold for $x' \neq \bar{l}$, for either specification of $V(l, \mu)$.

The argument for $I_1(\bar{l}, l')$ is analogous. □

The following two Lemmas establish that $C(\bar{l}, l')$ is differentiable on $(\bar{l}, 1)$ except at a finite number of points.

Lemma A.10. *The set of points in $] \bar{l}, 1[$ not contained in a fixed-Z interval is finite.*

Proof. Assume that some point $x \in] \bar{l}, 1[$ does not meet the criteria used in the definition of a fixed-Z interval. That is, there exists some $\bar{\mu} \in [0, 1]$ such that $(x, \bar{\mu}) \notin E$ and that either the left or right limit at $\bar{\mu}$ of $V(\bar{l}, \mu) - V(x, \mu)$ is 0. I will refer to the former requirement as the **boundary condition** on $(x, \bar{\mu})$ and the latter as the **limit indifference condition**.

I will show that there are only a finite number of points $x \in (\bar{l}, 1)$ at which a $\bar{\mu}$ exists such that $(x, \bar{\mu})$ satisfies both the boundary and limit indifference conditions.

The boundary condition implies that one of the following holds:

- $\bar{\mu} = 0$
- $\bar{\mu} = 0.5$
- $\bar{\mu} = 1$

- $(x, \bar{\mu}) \in X_1 \cup X_2$

In the first three cases it follows straightforwardly that the limit indifference condition holds only at a finite set of values of x ; substituting each of these three values of $\bar{\mu}$ into the condition gives an equation for which the solutions are a subset of the roots of a polynomial.

In the case of $(x, \bar{\mu}) \in X_1 \cup X_2$, note that it follows from Propositions 3 and 8 that for a given $\bar{\mu}$, there is at most one $x \in]\bar{l}, 1[$ such that $(x, \bar{\mu})$ satisfies the limit indifference condition. Solving for this x and substituting into the conditions for $(x, \bar{\mu}) \in X_1$ (posterior belief resulting from the less-likely report must be equal to 0.5) or $(x, \bar{\mu}) \in X_2$ (must have $\frac{d}{d\mu}V(\bar{l}, \bar{\mu}) - \frac{d}{d\mu}V(x, \bar{\mu}) = 0$) yields in each case an equation in $\bar{\mu}$ with solutions that are subsets of roots of a polynomial.

Since there are a finite number of solutions in each case, the set of points $(x, \bar{\mu})$ that meet both conditions is finite. \square

Lemma A.11. *$C(\bar{l}, l')$ is differentiable on any fixed-Z interval.*

Proof. Lemmas A.6 through A.9 establish that on a fixed-Z interval, the only elements of the specification of $C(\bar{l}, l')$ that vary with l' are terms of the form $F(z_i(l'))$ (the coefficients of these terms remain fixed over the interval since I_0 is constant and the ordering of the elements of Z is maintained). Since F is differentiable and each z function is differentiable, $C(\bar{l}, l')$ is differentiable on a fixed-Z interval. \square

At the finite number of points in $]\bar{l}, 1[$ not contained in fixed-Z intervals, $C(\bar{l}, l')$ may not be differentiable, but I will show that it is at least continuous. I will argue for left-continuity; the argument for right-continuity is analogous.

It will be useful to consider an arbitrary point $x \in]\bar{l}, 1[$ that is not contained in a fixed-Z interval. As there are a finite number of points x which are boundary points of a z function and therefore not contained in a fixed-Z interval, it follows that each such point is a boundary point of fixed-Z intervals; that is, there exist $a < x$ and $b > x$ such that $]a, x[$ and $]x, b[$ are fixed-Z intervals. In the following arguments I will use a' to denote an arbitrary point in the interior of a fixed-Z interval $]a, x[$, and I will use the following notation in order to refer easily to the limit at x of a z function defined on $]a, x[$:

$$z_{xi} = \lim_{l' \uparrow x} z_i(l'; \bar{l}, a')$$

Lemma A.12. *For each $Z_i(\bar{l}, x) \in Z(\bar{l}, x)$, $\exists j$ such that $\lim_{l' \uparrow x} Z_j(\bar{l}, l') = Z_i(\bar{l}, x)$*

Proof. Assume without loss of generality that $\frac{d}{d\mu}V(\bar{l}, Z_i(\bar{l}, x)) - \frac{d}{d\mu}V(x, Z_i(\bar{l}, x)) < 0$. That is, that the sign of $V(\bar{l}, \mu) - V(x, \mu)$ changes from positive to negative at $Z_i(\bar{l}, x)$. Assume

also without loss of generality that on some small fixed- Z interval with x as its upper bound, say $]x - \delta, x[$, $V(\bar{l}, Z_i(\bar{l}, x)) - V(l', Z_i(\bar{l}, x)) < 0$ ¹⁴. This implies that for any $\hat{\mu}$ less than but sufficiently close to $Z_i(\bar{l}, x)$, there exists an $x' \in]x - \delta, x[$ such that $V(\bar{l}, \hat{\mu}) - V(x', \hat{\mu}) < 0$ and $V(\bar{l}, \hat{\mu}) - V(x, \hat{\mu}) > 0$. Let Z_j be the closest sign change point for $l' = x - \delta$ that is below $Z_i(\bar{l}, x)$; that is:

$$j = \max \{n \in |Z(\bar{l}, x - \delta)| : Z_n(\bar{l}, x - \delta) < Z_i(\bar{l}, x)\}$$

By the continuity of $V(\bar{l}, \mu) - V(l', \mu)$, for any $\hat{\mu}$ less than but sufficiently close to $Z_i(\bar{l}, x)$, there exists an ϵ such that:

$$V(\bar{l}, \hat{\mu}) - V(l', \hat{\mu}) > 0, \forall l' \in]x - \epsilon, x[$$

It follows that $\forall l' \in]x - \epsilon, x[$, $Z_j(\bar{l}, l') \in]\hat{\mu}, Z_i(\bar{l}, x)[$. Therefore:

$$\lim_{l' \uparrow x} Z_j(\bar{l}, l') = Z_i(\bar{l}, x)$$

□

Note that if $Z_i(\bar{l}, x) \neq 0.5$, Lemma A.12 implies that there exists a function $z_j \in \hat{Z}(\bar{l}, a')$ such that:

$$z_{xj} = Z_i(\bar{l}, x)$$

If $Z_i(\bar{l}, x) = 0.5$, then the Lemma implies either the existence of a z_j function with $z_{xj} = 0.5$, or that $0.5 \in Z(\bar{l}, a')$.

Lemma A.13. *If there exists a function $z_i \in \hat{Z}(\bar{l}, a')$ such that $x = \omega_+(z_i)$ and z_{xi} is not equal to 0, 0.5, 1, or a local extremum of $V(\bar{l}, \mu) - V(x, \mu)$, then $z_{xi} \in Z(\bar{l}, x)$.*

Proof. For any $l' \in]a, x[$, $V(\bar{l}, z_i(l')) - V(l', z_i(l')) = 0$. It follows from the continuity of $V(\bar{l}, \mu) - V(x, \mu)$ away from $\mu = 0.5$ that:

$$\begin{aligned} \lim_{l' \uparrow x} V(\bar{l}, z_i(l')) - V(l', z_i(l')) &= 0 \\ V(\bar{l}, z_{xi}) - V(x, z_{xi}) &= 0 \end{aligned}$$

Since z_{xi} is a zero of $V(\bar{l}, \mu) - V(x, \mu)$, is in the interior of that function's domain, and by assumption is not a local extremum of the function or at a point of discontinuity, it must be a sign change point. □

Lemma A.14. *If $\bar{\mu}$ is a local extremum of $V(\bar{l}, \mu) - V(x, \mu)$, then the following set has an even number of members:*

$$\{z_i \in \hat{Z}(\bar{l}, a') : z_{xi} = \bar{\mu}\}$$

That is, if x is the upper bound of the domain of some z function because the function approaches a local extremum point at x , then there are an even number of z functions that approach that local extremum point at x .

¹⁴The argument is analogous if the function is positive on $]x - \delta, x[$. If it is constantly zero on $]x - \delta, x[$ then the argument is trivial.

Proof. Assume without loss of generality that there exists a function z_i with $z_{xi} = \bar{\mu}$ and that $\frac{d}{d\mu}V(\bar{l}, z_i(l')) - \frac{d}{d\mu}V(l', z_i(l')) < 0$ for any $l' \in \omega(z_i)$ (recall that the sign of this derivative is either always negative or always positive for a given z function); that is, z_i gives a positive-to-negative sign-change point. Assume further that (x, z_{xi}) is a local minimum of $V(\bar{l}, \mu) - V(l', \mu)$. Clearly $z_{xi} \notin Z(\bar{l}, x)$, as a local extremum with no discontinuity cannot be a sign change point. Consider the following (potential) member of $Z(\bar{l}, x)$:

$$Z_{next} = \min \{Z \in Z(\bar{l}, x) : Z > z_{xi}\}$$

By Lemma A.12, if Z_{next} exists, it is the limit of some $z_j \in \hat{Z}(\bar{l}, a')$, $j > i$. Since z_{xi} is a local minimum, Z_{next} must give a positive-to-negative sign change. Since z_{i+1} , if it exists, must always give a negative-to-positive sign change, $z_j \neq z_{i+1}(l'; \bar{l}, a')$, which implies that $z_{x(i+1)} \notin Z(\bar{l}, x)$ and $z_{x(i+1)} < Z_{next}$. Now assume that z_{i+1} exists and that $z_{xi} \neq z_{x(i+1)}$, and consider the following identity for some $y > 1$, which follows from the continuity of V away from $\mu = 0.5$

$$\lim_{l' \uparrow x} V(\bar{l}, \frac{z_i(l') + z_{i+1}(l')}{y}) - V(l', \frac{z_i(l') + z_{i+1}(l')}{y}) = V(\bar{l}, \frac{z_{xi} + z_{x(i+1)}}{y}) - V(x, \frac{z_{xi} + z_{x(i+1)}}{y})$$

Since $\frac{z_{xi} + z_{x(i+1)}}{y} \in]z_{xi}, Z_{next}[$, the right hand side of the above equation is weakly positive. As $V(\bar{l}, \mu) - V(x, \mu)$ is not constantly zero on any interval, there exists some y for which it is strictly positive. Since the expression inside the limit on the left hand side of the equation is weakly negative for any $l' \in]a, x[$, the limit is weakly negative for any y . This is a contradiction, so $z_{xi} = z_{x(i+1)}$ if z_{i+1} exists. If z_{i+1} does not exist, then either z_i is the last member of $\hat{Z}(\bar{l}, a')$ or $Z_{i+1}(\bar{l}, a') = 0.5$. In either case a similar argument applies to demonstrate a contradiction; for example, if z_i is the last member of $\hat{Z}(\bar{l}, a')$ then the following identity can be used:

$$\lim_{l' \uparrow x} V(\bar{l}, \frac{z_i(l') + 1}{y}) - V(l', \frac{z_i(l') + 1}{y}) = V(\bar{l}, \frac{z_{xi} + 1}{y}) - V(x, \frac{z_{xi} + 1}{y})$$

As before, the right-hand side must be strictly positive for some y while the left-hand side is weakly negative. For the other case, the same argument applies to the analogous expression using the point $\frac{z_{xi} + 0.5}{y}$.

These arguments demonstrate that whenever a z function that gives a positive-to-negative sign change point approaches a local extremum, so does a consecutive z function that gives a negative-to-positive sign change point. If more than two functions approach the same local extremum, assuming that the total number that do so is odd and following an analogous argument using the first and last of those functions yields a contradiction in the same way.

Analogous arguments apply using $z_{x(i-1)}$ when (x, z_{xi}) is a local maximum of $V(\bar{l}, \mu) - V(l', \mu)$, or if it is a local minimum but z_i gives a negative-to-positive sign change point. \square

Lemma A.15. $0.5 \in Z(\bar{l}, a') \cap 0.5 \notin Z(\bar{l}, x)$ if and only if the following set has an odd number of members:

$$\{z_i \in \hat{Z}(\bar{l}, a') : z_{xi} = 0.5\}$$

Proof. Let $0.5 = Z_j(\bar{l}, a')$ and assume without loss of generality that it is a positive-to-negative sign change point. Assume also that $0.5 \notin Z(\bar{l}, x)$ because the sign of $V(\bar{l}, \mu) - V(x, \mu)$ is weakly positive on an interval $(0.5 - \epsilon, 0.5 + \epsilon)$, and finally that $z_{x(j+1)} \neq 0.5$. Consider the following identity, which holds if z_{j+1} exists:

$$\lim_{l' \uparrow x} V(\bar{l}, \frac{0.5+z_{j+1}(l')}{y}) - V(l', \frac{0.5+z_{j+1}(l')}{y}) = V(\bar{l}, \frac{0.5+z_{x(j+1)}}{y}) - V(x, \frac{0.5+z_{x(j+1)}}{y})$$

Given the assumptions, the expression inside the limit on the left-hand side is weakly negative for any $l' \in]a, x[$, and for some y the right-hand side is strictly positive - a contradiction. If z_{j+1} does not exist, then the same argument applies for an analogous expression using any fixed μ between 0.5 and 1. For example:

$$\lim_{l' \uparrow x} V(\bar{l}, 0.75) - V(l', 0.75) = V(\bar{l}, 0.75) - V(x, 0.75)$$

To see that the number of z functions that approach 0.5 must be odd, note that if $z_{x(j+2)}$ was also equal to 0.5, the argument can then be repeated using $\mu = \frac{z_{x(j+2)} + z_{x(j+3)}}{y}$ (replacing $z_{x(j+3)}$ with 0 if it does not exist). Similarly, if $z_{x(j-1)} = 0.5$, use $\mu = \frac{z_{x(j-1)} + z_{x(j-2)}}{y}$.

If $V(\bar{l}, \mu) - V(x, \mu)$ is weakly negative in an interval around x , an analogous argument applies beginning with z_{j-1} instead of z_{j+1} .

In the converse direction, assume again that $0.5 = Z_j(\bar{l}, a')$ and that it is a positive-to-negative sign change point. Assume also that an odd number of z functions approach 0.5 at x . Additionally, assume without loss of generality that an even number of z functions approach 0.5 from below, and an odd number from above. Let Z_{prev} be the largest member of $Z(\bar{l}, x)$ that is less than 0.5, if one exists, and let Z_{next} be the smallest member of $Z(\bar{l}, x)$ that is greater than 0.5, if one exists. If Z_{prev} exists, it is equal to $z_{x(j-1-m)}$, with m even. To see this, note that m is the number of z functions below 0.5 that either approach 0.5 at x (an even number by assumption) or approach a local extremum of $V(\bar{l}, \mu) - V(x, \mu)$ between Z_{prev} and 0.5 (shown to be even above). This implies that Z_{prev} is a negative-to-positive sign change point and that $V(\bar{l}, \mu) - V(x, \mu)$ is weakly positive on some interval $]0.5 - \epsilon, 0.5[$. The same logic applied above 0.5 shows that Z_{next} , if it exists, is a negative-to-positive sign change point and that $V(\bar{l}, \mu) - V(x, \mu)$ is weakly positive on some interval $]0.5, 0.5 + \epsilon[$. Therefore, on the assumption that Z_{prev} and Z_{next} both exist, $0.5 \notin Z(\bar{l}, x)$.

If Z_{prev} does not exist, then all z_i functions in $\hat{Z}(\bar{l}, a')$ with $i < j$ must either approach 0, 0.5, or a local extremum of $V(\bar{l}, \mu) - V(x, \mu)$ at x . Assume without loss of generality that there are an even number of such functions (that j is odd). This implies that $I_0(\bar{l}, a') = 1$. Since, as argued above, an even number of z functions approach 0.5 from below at x or approach a local extremum at x , an even number of functions must approach 0. Let z_h be the largest of those functions. That is:

$$h = \max\{i \in |\hat{Z}(\bar{l}, a')| : z_{xi} = 0\}$$

Since h is even and $I_0(\bar{l}, a') = 1$, z_h gives a negative-to-positive sign change point. Now consider the following identity:

$$\lim_{l' \uparrow x} V(\bar{l}, \frac{z_h(l') + z_{h+1}(l')}{y}) - V(l', \frac{z_h(l') + z_{h+1}(l')}{y}) = V(\bar{l}, \frac{z_{x(h+1)}}{y}) - V(x, \frac{z_{x(h+1)}}{y})$$

Since z_h gives a negative-to-positive sign change point, the expression inside the limit on the left hand side is weakly positive for any $l' \in]a, x[$. Therefore the expression on the right-hand side is weakly positive, and since there are no members of $Z(\bar{l}, x)$ that are less than 0.5, $V(\bar{l}, \mu) - V(x, \mu)$ is weakly positive on $]0, 0.5[$, and the argument is the same as in the case where Z_{prev} exists. An analogous argument applies if Z_{next} does not exist. \square

Lemma A.16. $I_0(\bar{l}, a') \neq I_0(\bar{l}, x)$ if and only if the following set has an odd number of members:

$$\{z_i \in \hat{Z}(\bar{l}, a') \mid z_{xi} = 0\}$$

Lemma A.17. $I_1(\bar{l}, a') \neq I_1(\bar{l}, x)$ if and only if the following set has an odd number of members:

$$\{z_i \in \hat{Z}(\bar{l}, a') \mid z_{xi} = 1\}$$

The proofs of lemmas A.16 and A.17 are analogous to that of A.15; note that, for example, $I_0(\bar{l}, a') = 1$ and $I_0(\bar{l}, a') = 0$ implies that $Z_1(\bar{l}, a')$ is a positive-to-negative sign change point and $Z_1(\bar{l}, x)$ a negative-to-positive sign change point.

Lemma A.18. $C(\bar{l}, l')$ is continuous at x .

Proof. $C(\bar{l}, l')$ is left-continuous at x if the following holds:

$$\begin{aligned} I_1(\bar{l}, a') &+ \lim_{l' \uparrow x} \sum_{Z_i \in Z(\bar{l}, l')} [(-1)^{i-I_0(\bar{l}, a')} F(Z_i)] \\ &= I_1(\bar{l}, x) + \sum_{Z_i \in Z(\bar{l}, x)} [(-1)^{i-I_0(\bar{l}, x)} F(Z_i)] \end{aligned}$$

Note that the limit of the summation on the left-hand-side above can be rewritten as a sum of individual limits. If $0.5 \notin Z(\bar{l}, a')$, then:

$$\lim_{l' \uparrow x} \sum_{Z_i \in Z(\bar{l}, l')} [(-1)^{i-I_0(\bar{l}, a')} F(Z_i)] = \sum_{z_i \in \hat{Z}(\bar{l}, a')} [(-1)^{i-I_0(\bar{l}, a')} F(z_{xi})]$$

Alternately, if $0.5 = Z_j(\bar{l}, a')$, then:

$$\begin{aligned} \lim_{l' \uparrow x} \sum_{Z_i \in Z(\bar{l}, l')} [(-1)^{i-I_0(\bar{l}, a')} F(Z_i)] &= \sum_{z_i \in \hat{Z}(\bar{l}, a') \setminus \{0.5\}} [(-1)^{i-I_0(\bar{l}, a')} F(z_{xi})] \\ &\quad + (-1)^{j-I_0(\bar{l}, a')} F(0.5) \end{aligned}$$

Lemmas A.12 through A.17 demonstrate that, since F is continuous, for each $Z_i \in Z(\bar{l}, x)$, there exists a $Z_j \in Z(\bar{l}, a')$ such that:

$$\lim_{l' \uparrow x} (-1)^{j-I_0(\bar{l}, a')} F(Z_j(\bar{l}, l')) = (-1)^{i-I_0(\bar{l}, x)} F(Z_i(\bar{l}, x))$$

The equality of the coefficients follows from the fact that if i is even and j odd or vice versa, $I_0(\bar{l}, a') \neq I_0(\bar{l}, x)$ by Lemma A.16, as z_i functions with $i < j$ and $z_{xi} \notin Z(\bar{l}, x)$ must approach local extrema (an even number of functions - Lemma A.14), 0.5 (either an even number of functions or 0.5 is in exactly one of $Z(\bar{l}, a')$ and $Z(\bar{l}, x)$ - Lemma A.15), or 0.

Now consider those $z_i \in \hat{Z}(\bar{l}, a')$ for which $z_{xi} \notin Z(\bar{l}, x)$. Assuming that if $0.5 \in Z(\bar{l}, a')$ then $0.5 \in Z(\bar{l}, x)$, continuity holds if the sum of the expression $(-1)^{i-I_0(\bar{l}, a')} F(z_{xi})$ over these functions is equal to $I_1(\bar{l}, x) - I_1(\bar{l}, a')$. The sum of the expression over any even number of consecutive z_i functions with the same limit at x is 0, so the sum over all z_i functions that approach local extrema at x is 0 by Lemma A.14. $F(0) = 0$, so the sum over all z_i functions with $z_{xi} = 0$ is also 0. If an even number of z_i functions have $z_{xi} = 1$, the expression also sums to zero over those functions, and $I_0(\bar{l}, x) = I_1(\bar{l}, a')$, so continuity holds. If an odd number of functions have $z_{xi} = 1$, the expression sums to 0 over all but one of them; say the last one (highest i). Since $F(1) = 1$, the single remaining limit is equal to 1 if $I_1(\bar{l}, a') = 0$, and -1 if $I_1(\bar{l}, a') = 0$ ¹⁵. As an odd number of z functions approach 1 at x , this is equal to $I_1(\bar{l}, x) - I_1(\bar{l}, a')$, and left-continuity holds.

The remaining case is when $0.5 = Z_j(\bar{l}, a')$ and $0.5 \notin Z(\bar{l}, x)$. In this case the argument is identical, save that by Lemma A.15 there must be an odd number of z_i functions with $z_{xi} = 0.5$. In the limit one of these cancels out the term $(-1)^{j-I_0(\bar{l}, a')} F(0.5)$, and the expression $(-1)^{i-I_0(\bar{l}, a')} F(z_{xi})$ sums to zero over the remaining even number of functions with $z_{xi} = 0.5$. Therefore left-continuity holds in every case.

It can be shown by symmetrical arguments that results analogous to Lemmas A.12 through A.17 hold for a fixed- Z interval $]x, b[$ and z functions with $\omega_-(z) = x$. Right-continuity follows straightforwardly. \square

Lemma A.19. *For a fixed \bar{l} , $C(\bar{l}, l')$ is continuous on $]\bar{l}, 1[$, and is differentiable on $]\bar{l}, 1[$ except at a finite number of points.*

Proof. Follows from Lemmas A.10, A.10, and A.18. \square

Lemma A.20. *For a fixed \bar{l} , $C(\bar{l}, l')$ is continuous on $(0, \bar{l})$, and is differentiable on $(0, \bar{l})$ except at a finite number of points.*

Proof. The arguments for Lemma A.19 also apply on the interval $(0, \bar{l})$. \square

With these characteristics of C established, the remainder of the proof largely follows the proof of Proposition 3 of Osborne and Pitchik (1986), which proves the analogous result for a one-dimensional location model in which the distribution of consumers is atomless, and consumers select the nearest firm. Some adaptations are required to handle the facts that C is not differentiable everywhere and that the distribution of consumer preferences may have an atom at 0.5.

¹⁵This holds because $i - I_0(\bar{l}, a')$ is even if $Z_i(\bar{l}, a')$ is a positive-to-negative sign change point, and odd if it is negative-to-positive

Let H^n be the CDF corresponding to the symmetric equilibrium mixed strategy with n firms. I first note that where G is continuous, so is H^n .

Lemma A.21. *For $n > 2$, the symmetric equilibrium strategy $H^n(l)$ is atomless except possibly at $l = 0.5$.*

Proof. The proof of Lemma 1 of Osborne and Pitchik (1986), which shows that any point in their strategy space is an atom of the equilibrium strategies of at most two firms, can be applied in this setting at a point $\bar{l} \in L$ if $C(\bar{l}, l')$ is continuous at $l' = \bar{l}$. By Lemma A.5, this holds for $\bar{l} \neq 0.5$. \square

Given Lemma A.21, the expected payoff of a firm locating at a point $\bar{l} \neq 0.5$ when the $n - 1$ other firms all follow a mixed strategy with CDF H^n , $P(l, H^n)$, is given by:

$$\begin{aligned} P(l, H^n) &= (n-1) \int_0^l (1 - C(l, u)) H^n(u)^{n-2} dH(u) \\ &\quad + (n-1) \int_l^1 (C(l, u))(1 - H^n(u))^{n-2} dH^n(u) \\ &\quad + \sum_{k=0}^{n-3} (n-1)(n-2) \binom{n-3}{k} \int_0^l \int_l^1 [(C(l, v) \\ &\quad - C(l, u)) H^n(u)^k (1 - H^n(v))^{n-k-3}] dH^n(v) dH^n(u) \end{aligned}$$

Note that since H^n is atomless away from 0.5, equilibrium requires that $P(l, H^n) = 1/n$ almost everywhere. As $P(l, H^n)$ is continuous in l , this implies that $P(l, H^n) = 1/n$ for all $l \in \text{supp } H^n$.

I define the following sets:

$$\underline{D}(l) = \{l' \in]0, l[: \nexists \frac{\partial}{\partial l'} C_2(l, l')\} \cap \{0, l\}$$

$$\bar{D}(l) = \{l' \in]l, 1[: \nexists \frac{\partial}{\partial l'} C_2(l, l')\} \cap \{l, 1\}$$

These sets are finite by Lemmas A.19 and A.20. Denote by $\underline{D}_i(l)$ and $\bar{D}_i(l)$ the i th element of the each set when ordered by size, i.e. $\underline{D}(l) = \{\underline{D}_1(l), \underline{D}_2(l) \dots \underline{D}_{\underline{m}}(l)\}$ with $0 = \underline{D}_1(l) < \underline{D}_2(l) \dots < \underline{D}_{\underline{m}}(l) = 1$. Splitting up the integrals in the expected payoff using the members of the D sets as limits, integrating by parts and applying the Binomial theorem gives:

$$\begin{aligned} P(l, H^n) &= C(l, 0)(1 - H^n(l))^{n-1} + (1 - C(l, 1))H^n(l)^{n-1} \\ &\quad + \sum_{i=1}^{\underline{m}-1} \int_{\underline{D}_i(l)}^{\underline{D}_{i+1}(l)} \frac{\partial C}{\partial l'}(l, u)(1 + H^n(u) - H^n(l))^{n-1} du \\ &\quad + \sum_{i=1}^{\bar{m}-1} \int_{\bar{D}_i(l)}^{\bar{D}_{i+1}(l)} \frac{\partial C}{\partial l'}(l, u)(1 + H^n(l) - H^n(u))^{n-1} du \end{aligned}$$

Combining the fact that $P(l, H^n) = 1/n$ for all $l \in \text{supp } H^n$ with Proposition 5 and the continuity of $nP(l, H^n)$ in l gives:

$$\lim_{n \rightarrow \infty} nP(l, H^n) = 1, \forall l \in \text{supp } G$$

Using the expression for $P(l, H^n)$:

$$\begin{aligned} nP(l, H^n) &= nC(l, 0)(1 - H^n(l))^{n-1} + n(1 - C(l, 1))H^n(l)^{n-1} \\ &+ \sum_{i=1}^{i=\underline{m}-1} \int_{\underline{D}_i(l)}^{\underline{D}_{i+1}(l)} n \frac{\partial C}{\partial l'}(l, u)(1 - (H^n(l) - H^n(u)))^{n-1} du \\ &+ \sum_{i=1}^{i=\bar{m}-1} \int_{\bar{D}_i(l)}^{\bar{D}_{i+1}(l)} n \frac{\partial C}{\partial l'}(l, u)(1 - (H^n(u) - H^n(l)))^{n-1} du \end{aligned}$$

If l is on the interior of the support of G , $0 < H(l) < 1$ by Proposition 5. It follows that there exists $\epsilon \in]0, 1[$ and \bar{n} such that for all $n > \bar{n}$, $H^n \in (\epsilon, 1 - \epsilon)$. Applying L'Hôpital's rule gives:

$$\begin{aligned} \lim_{n \rightarrow \infty} n(1 - \epsilon)^{n-1} &= 0 \\ \lim_{n \rightarrow \infty} n(\epsilon)^{n-1} &= 0 \end{aligned}$$

It follows that the first two terms in the above expression for $nP(l, H^n)$ go to 0 as $n \rightarrow \infty$.

For the integrals in the expression, I apply Lemma 2 of Osborne and Pitchik (1986). The Lemma states that for $c > 0$ and $n = 1, 2, \dots$, with functions $f^n : [0, c] \rightarrow [0, 1]$, $f : [0, c] \rightarrow [0, 1]$, $g : [0, c] \rightarrow \mathbb{R}$ that are nondecreasing and twice continuously differentiable, with $f^n(0) = 0$ and $f^n \rightarrow f$, $f^{n'} \rightarrow f'$, and $f^{n''} \rightarrow f''$ uniformly:

$$n \int_0^c g'(x)(1 - f^n(x))^{n-1} dx \rightarrow \begin{cases} \infty & f'(0) = 0, g'(0) > 0 \\ \frac{g'(0)}{f'(0)} & f'(0) > 0 \end{cases}$$

As part of their proof of this Lemma, Osborne and Pitchik also establish that for $0 < \epsilon < c$, if $f'(0) > 0$, then:

$$n \int_{\epsilon}^c g'(x)(1 - f^n(x))^{n-1} dx \rightarrow 0$$

Appropriate substitutions allow these results to be applied to each of the integrals in the two summations. For the integrals in the first summation, $c = \underline{D}_{i+1}(l)$, $x = l - u$, $g(x) = -C(l, l - x)$, and $f^n(x) = H^n(l) - H^n(l - x)$. For the integrals in the second summation, $c = \bar{D}_{i+1}(l) - z$, $x = u - l$, $g(x) = C(l, l + x)$, and $f^n(x) = H^n(l + x) - H^n(l)$. The first integral in each summation goes to $\frac{1}{h(l)} \frac{\partial C(l, l)}{\partial l'}$, and the rest go to zero. Since, as noted above, $nP(l, H^n)$ must go to 1, it follows that:

$$h(l) = 2 \frac{\partial C(l, l)}{\partial l'}$$

Lemma A.22. $\frac{\partial C(\bar{l}, \bar{l})}{\partial l} = \frac{\partial C(\bar{l}, \bar{l})}{\partial l'} = \frac{g(\bar{l})}{2}$

Proof. That $\frac{\partial C(\bar{l}, \bar{l})}{\partial l} = \frac{\partial C(\bar{l}, \bar{l})}{\partial l'}$ follows from the fact that $C(l, l') = C(l', l)$. Then, since $C(l, l) \equiv G(l)$:

$$g(\bar{l}) = \frac{\partial C(\bar{l}, \bar{l})}{\partial l} + \frac{\partial C(\bar{l}, \bar{l})}{\partial l'}$$

□

It follows that at any l in $\text{supp } G$ not equal to 0.5, $h(l) = g(l)$. Since H and G are equal at the upper and lower bounds of $\text{supp } G$, they are equal at any l not equal to 0.5. Since they are CDFs, they must also be equal at $l = 0.5$.

Proof of Proposition 7

Lemmas 6 and 7 imply that the unique consumer-optimal signal can be found by maximising the following function with respect to β :

$$E(U|(1, \beta)) = \mu + \beta(1 - \mu) + \rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}} \left[(\beta - \beta^2(1 - \mu))^{\frac{1}{2}} - \beta(1 - \mu)^{\frac{1}{2}} \right]$$

Since this function is concave by Lemma 7, it has a unique maximum either at the boundary of its domain (which is $[0, 1]$ since β is a probability) or when the following first-order condition is satisfied.

$$\rho(\mu - \frac{1}{2})\mu^{\frac{1}{2}}(1 - \mu)^{\frac{1}{2}} \left[(1 - \mu)^{\frac{1}{2}} - \frac{\frac{1}{2} - \beta(1 - \mu)}{(\beta - \beta^2(1 - \mu))^{\frac{1}{2}}} \right] = 1 - \mu$$

That this condition is satisfied by $\hat{\beta}(\mu; \rho)$ can be verified by substitution. The concavity of $E(U|(1, \beta))$ implies that $\beta = 1$ is optimal where $\hat{\beta}(\mu; \rho)$ is greater than 1.

Chapter 2

How Informed Should Voters Be?

2.1 Introduction

When decisions are made through voting, the voters are typically imperfectly informed about the choices they are voting on. Faced with propaganda, fake news, or simply a lack of information about the potential effects of voting outcomes, it is natural to consider the potential benefits of providing voters with more accurate information. A social planner may be able to improve the accuracy of voters' information through a variety of channels, such as direct information provision, active research into outcomes, or regulation of media organisations.

However, if this social planner is utilitarian, voting carries its own problems. Even when much-discussed issues of strategic voting are avoided by limiting a ballot to a binary choice, the fact that the decision rule does not make use of information about voters' cardinal preferences¹ leads to 'tyranny of the majority' effects, in which the weak preferences of a majority outweigh the strong preferences of a minority, to a point at which total utility would be increased by reversing the decision. A considerable literature addresses potential solutions to this problem that can be applied over multiple votes, whether by allowing voters to trade votes (Buchanan and Tullock (1962), Casella and Palfrey (2019)), allowing them to store votes to use on future issues (Casella (2005), Casella and Gelman (2008)), charging them appropriately-structured costs for increasing numbers of votes (Lalley and Weyl (2018)), or allowing them to abstain when voting is costly (Borgers (2004)).

In this paper, I show that restricting voter information in certain ways can also address these problems even when there is only a single vote (or when voters cannot be tracked across votes), and potentially increase total expected utility, demonstrating that it is not necessarily the case that the better informed the voters are, the better the outcome. When

¹In fact it cannot make use of them if the outcome is binary, valuations are independent, and incentive-compatibility is required.

voters have imperfect information about their valuations of the choices on a ballot, they sometimes make ‘mistakes’, voting for a particular option when they would prefer another if they knew their valuation with certainty. When the information structure is such that mistakes are primarily or exclusively made by relatively-ambivalent voters, the occurrence of ‘tyranny of the majority’ outcomes can be reduced relative to the case in which voters are perfectly informed. On the other hand, when voters make mistakes, the quality of the information about preferences provided to the decision-making mechanism, the vote, declines. This is the trade-off faced by a utilitarian social planner who can influence the information available to voters.

The model of this paper involves a binary vote over two outcomes, using simple majority rule. Voters have a shared, commonly-known valuation of one of the outcomes (the ‘status quo’). Their valuations of the other outcome are independent and identically distributed. A social planner selects the information structure through which voters learn about their own valuations. This information structure consists of a stochastic mapping from a voter’s valuation to a probability distribution on a set of messages; the voters observe the choice of information structure, and update their beliefs rationally on receiving a message.

The planner in this model faces some fundamental restrictions, in that she is limited to information structures that condition the messages a voter receives on that voter’s *own* valuation. That is, she cannot send messages that only provide aggregate information about valuations. If she could, it would be straightforward to implement first-best in this setting, by sending messages that simply indicated to every voter which outcome would maximise total utility.

The assumption that this is not possible, as well as the assumption that voter valuations are independent, is intended to situate this model as one in which voters ‘care about different things’. In reality, a given agent’s valuation of some possible voting outcome depends on both the effects of the outcome, and the agent’s preferences over those effects. As the range of possible preferences agents may have becomes more diversely distributed, information about the effects of an outcome becomes less and less informative about the valuation of a randomly chosen agent, if the preferences of that agent are unknown. A planner lacking knowledge of those preferences can still control how the agent learns about their own valuation, by controlling the information they receive about the effects of the policy. But this does not give the planner direct access to the valuations themselves, so she cannot aggregate them. To do so, she must elicit information about preferences from the agents, which is what the vote does!

The model of this paper captures such a scenario by applying the restrictions described above to the planner. Subject to these restrictions, I otherwise grant the planner extensive powers in choosing the information structure. This does allow the use of complex information structures unlikely to be plausible in practice. However, this is of limited concern given my results. The optimal structures I derive are always simple partitions of the space of valuations, with a small number of possible messages.

When abstentions are prohibited, I show that perfect information is optimal if and only if a

simple symmetry condition on the distribution of valuations holds: that in expectation, the degree to which voters for one policy prefer it to the other is equal to the degree to which voters in the reverse direction prefer the reverse outcome. In other words, the expected benefit to voters for a given outcome of their chosen outcome is equalised across the two outcomes. If this condition holds, ‘tyranny of the majority’ outcomes still occur, but restricting voter information cannot reduce them: reducing their incidence in favour of one outcome is always offset by increasing it in the other. Otherwise the optimal information structure induces some of the most ambivalent voters for one of the options to vote incorrectly. This reduces the difference between the expected benefit of winning for voters on the two sides. If an information structure exists which eliminates this difference altogether (existence depends on whether the space of valuations is large enough), then the optimal information structure approaches it as the number of voters grows large. These results mirror the emphasis on expected benefits of winning a vote in the literature on optimal binary voting mechanisms; in Azrieli and Kim (2014), optimal voting rules weight each voter by the benefit they derive from their preferred outcome winning.

If indifferent voters abstain from the vote, then the planner gains more power, since she can design an information structure that sends some voters indifference-inducing messages, resulting in abstention². In this case, restriction of information can be beneficial even when the symmetry condition is met and perfect information would be optimal in the no-abstentions case. I show that for symmetrical valuation distributions with abstentions, perfect information is never optimal if the number of voters is large enough. The optimal information structures here induce the most ambivalent voters to abstain, providing an effect on participation that is essentially equivalent to adding a cost to voting, but may be possible when the imposition of a cost is not (in the case of monetary costs, variation in wealth may make the imposition of uniform costs impossible, or social and political obstacles may prevent the use of such mechanisms).

The planner’s commitment to an information structure which is observed by the voters connects this model to the information design literature, in which a sender commits to an information structure, observable to a receiver, that ‘persuades’ the receiver in a way that is beneficial to the sender. Building from the single-sender single-receiver model of Kamenica and Gentzkow (2008), this literature includes several voting models. Closest to the model of this paper is that of Van Straeten and Yamashita (2019), which similarly considers a utilitarian sender, but does not restrict her to information structures that condition on a voter’s own valuation only. They include an additive private component to voters’ valuations, which prevents the sender straightforwardly implementing first-best. They show that in many cases a signal structure that simply reports the average of the accessible component of the valuations is still optimal despite the presence of the private component.

Other voting papers in the information design literature typically focus on a sender who wishes to persuade voters to select a particular option. For example, Bardhi and Guo (2017) analyze the case of a sender who wishes to persuade a group of voters to take a certain

²Inducing such a precise belief may be challenging in reality, but approximate indifference is sufficient if voting carries small inherent costs.

decision, using private signals, when the voters are ex-ante heterogeneous in their ‘thresholds of doubt’ (how much convincing they require to approve the decision). Similarly, Alonso and Camara (2016) also consider how a sender can influence a vote towards a particular outcome, in this case using a public signal and majority rule with varying thresholds. The model of this paper differs in that the sender (the planner) is motivated by the interests of the voters in aggregate. This creates a motive for the information structure to be informative, since the planner cares about the preferences of voters and has a need for them to be informed about their preferences in order to make use of that information, rather than just to manipulate their beliefs. In fact, the planner in this model is something of a redundant figure. Since the voters are ex ante identical, the optimal information structures I derive are ones that the voters would agree to commit to ex ante, if they are expected utility maximisers.

Sections 2.2.1 and 2.2.2 introduce the structure of the baseline model with mandatory voting and describe the planner’s problem. Section 2.2.3 presents results on optimal information structures, which Section 2.2.4 generalises to voting mechanisms with differing thresholds. Section 2.3 considers non-mandatory voting, allowing for abstentions. Section 2.4 concludes.

2.2 Baseline Model with Mandatory Voting

2.2.1 Structure and Timing

A social planner determines the information structure facing Q agents, Q odd, who will vote on whether to implement a policy. The outcome is binary: either the policy is implemented (outcome Y) or it is not (outcome N). If the outcome is N , each agent receives an identical commonly-known payoff, v . I assume without loss of generality that $v \geq 0$. Agent’s payoffs if the outcome is Y vary across agents and are uncertain ex-ante; I denote agent i ’s utility under Y by θ_i . Values of θ_i are drawn IID from a continuous distribution on a closed interval $\Theta = [\underline{\theta}, \bar{\theta}]$, $v \in (\underline{\theta}, \bar{\theta})$, with a differentiable and strictly increasing cumulative distribution function F . $E(\theta_i)$ is normalised to 0.

The timing of the model is as follows:

- The social planner publicly commits to a **signal structure**.
- The agents receive private messages and update their beliefs about their payoffs.
- The agents vote on whether to implement the policy.
- The result of the vote is implemented and agents receive their payoffs.

The social planner determines how much information the agents will receive about their own payoffs before they vote. Specifically, she selects a signal structure, which is a set $S = \{M, G\}$,

where M is a finite set of messages and $G : \Theta \rightarrow \Delta M$ an F -measurable function that assigns a probability distribution over messages to each possible true valuation θ . I use g to denote the resulting conditional probabilities of messages given valuations. That is, for any message $m \in M$ and valuation $\theta \in \Theta$ ³:

$$g(m|\theta) = G(\theta)(m)$$

Given G , applying Bayes' rule on receipt of a particular message m generates a posterior distribution on θ , with a cumulative distribution function H given by:

$$H(\theta|m) = \frac{\int_{\underline{\theta}}^{\theta} g(m|t)f(t) dt}{\int_{\underline{\theta}}^{\bar{\theta}} g(m|t)f(t) dt}$$

It follows that, given G , any message m generates a posterior expectation of θ :

$$E(\theta|m) = \int_{\underline{\theta}}^{\bar{\theta}} \theta dH(\theta)$$

As discussed in the introduction, the information structure the planner chooses maps from a *single* valuation to probabilities over messages. She cannot condition the messages a voter receives on the valuations of other voters.

The planner's choice of signal structure is public and made with full commitment - agents are fully aware of the properties of the signal structure she selects, and she cannot alter it or distort the information provided at any later stage.

The planner's objective is to maximise the ex ante expected sum of agent utilities. Since agents are ex ante identical, it is equivalent to treat the planner as maximising the ex ante expected utility of a single agent. I will denote this expectation by $\bar{U}(S)$:

$$\bar{U}(S) = v + \sum_{m \in M} (E(\theta|m) - v)P(Y|m)P(m)$$

Once agents receive a message, they update their beliefs about their valuation θ_i according to Bayes' rule and the properties of the signal structure S . They then take an action $a_i \in \{y, n\}$ (their vote). I use $P(Y)$ to denote the probability that outcome Y wins the vote under signal structure; that is, the probability that greater than $\frac{Q-1}{2}$ voters receive messages that induce them to vote y . Consequently, $P(Y|m)$ denotes the probability that outcome Y wins conditional on a single voter having received message m (and voted accordingly). The vote induced by a message m follows straightforwardly from $E(\theta|m)$, as I will describe in detail below.

Note that this setup assumes that voting is mandatory, an assumption I will revisit in Section 2.3. The policy outcome is determined by majority vote. Agents choose their action

³Where I refer to multiple signal structures in later results and proofs, I follow this lowercase convention. For example, if the function assigning distributions over messages is G' , g' denotes the resulting conditional probabilities, and so on.

to maximise their expected utility given the beliefs induced by the message they received (randomizing 50-50 if indifferent).

2.2.2 The Planner's Problem

Given that the vote is binary, agents' strategies are relatively trivial; it is the planner's choice of information structure that is of primary interest. Denoting the message received by agent i by m_i , an agent's action is simply determined by the posterior expectation of θ_i given the message m_i .

$$P(a_i = y|m_i) = \begin{cases} 0 & E(\theta_i|m_i) < v \\ 0.5 & E(\theta_i|m_i) = v \\ 1 & E(\theta_i|m_i) > v \end{cases}$$

The fact that votes are entirely determined by the sign of $E(\theta_i|m_i) - v$ simplifies the planner's problem, as it provides a straightforward relationship between the posterior beliefs induced by messages and the resulting votes. Proposition 1 demonstrates that since there are only two actions, then by a revelation principle-type argument I can focus on signal structures with only two messages in which the messages are action recommendations.

Proposition 1. *Let $S = \{M, G\}$ be a signal structure, with $|M| > 2$. Then there exists a signal structure $S' = \{M', G'\}$ such that $|M'| = 2$ and $\bar{U}(S') = \bar{U}(S)$.*

Proof. First note that since messages carry no information about the actions of other agents, then if two messages m and m' induce the same action, it follows that $P(Y|m) = P(Y|m')$.

Divide M into three subsets as follows:

$$M_y = \{m \in M : E(\theta|m) > v\}$$

$$M_n = \{m \in M : E(\theta|m) < v\}$$

$$M_d = \{m \in M : E(\theta|m) = v\}$$

All messages in M_y induce a y vote, all messages in M_n induce an n vote, and all messages in M_d induce randomization. Therefore, using the fact that $E(\theta|m \in M_d) - v = 0$, $\bar{U}(S)$ can be written as:

$$\begin{aligned} \bar{U}(S) &= v + P(Y|a = y, S)P(m \in M_y)(E(\theta|m \in M_y) - v) \\ &\quad + P(Y|a = n, S)P(m \in M_n)(E(\theta|m \in M_n) - v) \end{aligned}$$

Construct S' as follows:

$$M' = \{y, n\}$$

$$g'(y|\theta) = \sum_{m \in M_y} g(m|\theta) + \frac{1}{2} \sum_{m \in M_d} g(m|\theta)$$

$$g'(n|\theta) = \sum_{m \in M_n} g(m|\theta) + \frac{1}{2} \sum_{m \in M_d} g(m|\theta)$$

This gives:

$$P(m' = y) = P(m \in M_y) + \frac{1}{2}P(m \in M_d)$$

$$P(m' = n) = P(m \in M_n) + \frac{1}{2}P(m \in M_d)$$

$$(E(\theta|m' = y) - v)P(m' = y) = (E(\theta|m \in M_y) - v)P(m \in M_y)$$

$$(E(\theta|m' = n) - v)P(m' = n) = (E(\theta|m \in M_n) - v)P(m \in M_n)$$

Clearly the probability that any other voter votes y is the same under S' as under S , and it follows that $P(Y|a = y, S') = P(Y|a = y, S)$. Therefore:

$$\begin{aligned} \bar{U}(S') &= v + P(Y|a = y, S')P(m' = y)(E(\theta|m' = y) - v) \\ &\quad + P(Y|a = n, S')P(m' = n)(E(\theta|m' = n) - v) \\ &= v + P(Y|a = y, S)P(m \in M_y)(E(\theta|m \in M_y) - v) \\ &\quad + P(Y|a = n, S)P(m \in M_n)(E(\theta|m \in M_n) - v) \\ &= \bar{U}(S) \end{aligned}$$

□

I will restrict attention from this point on to binary signal structures in which $M = \{y, n\}$, $E(\theta_i|m_i = y) \geq v$, and $E(\theta_i|m_i = n) \leq v$. That is, structures in which the messages are action recommendations and it is incentive compatible for voters to follow the recommendations they receive. For brevity, I will denote $E(\theta_i|m_i = m)$ under a signal structure S by $\hat{\theta}_S(m)$, and the probability of an event X under a signal structure S by $P_S(X)$. In these probabilities references to y and m may be taken as denoting either the message or the resulting action (since for these information structures receiving a message always results in taking the corresponding action). Probabilities conditional on a message/action are given for

a single agent; for example, $P_S(Y|y)$ is the probability that the outcome is Y given that one particular agent votes y .

Using the fact that for any S , $\hat{\theta}_S(y)P_S(y) + \hat{\theta}_S(n)P_S(n) = 0$, it will be useful to rewrite \bar{U} as follows:

$$\begin{aligned}\bar{U}(S) &= v + P_S(Y|y)P_S(y)(\hat{\theta}_S(y) - v) + P_S(Y|n)P_S(n)(\hat{\theta}_S(n) - v) \\ &= vP_S(N) + (P_S(Y|y) - P_S(Y|n))P_S(y)\hat{\theta}_S(y)\end{aligned}$$

This reformulation breaks down an agent's expected utility into two parts corresponding to the two outcomes. The first term straightforwardly captures the known payoff of v for the N outcome. The second term expresses the potential for benefiting from a Y outcome. If the agent could not influence the outcome, this would be 0, since $E(\theta) = 0$. However, since the agent gets a vote, the probability of Y depends on the agent's payoff. The term $(P_S(Y|y) - P_S(Y|n))P_S(y)$ gives the probability that the agent votes y and in doing so changes the outcome of the vote (her vote is pivotal). The $\hat{\theta}_S(y)$ term gives the expected benefit to the agent of affecting the outcome in this way.

I will denote the impact of an agent's vote on $P(Y)$, given S , by $\rho(S)$:

$$\rho(S) = P_S(Y|y) - P_S(Y|n)$$

This gives:

$$\bar{U}(S) = vP_S(N) + \rho(S)P_S(y)\hat{\theta}_S(y)$$

The above formulation of $\bar{U}(S)$ implies that for a fixed $P(y)$, \bar{U} is increasing in $\hat{\theta}(y)$. This leads to Lemma 1, which implies that optimality can be achieved within an even simpler class of signal structures, that of **monotone partitions**. A binary signal structure with $M = \{y, n\}$ is a monotone partition if there exists a cut-off x such that n is sent whenever θ is below x and y is sent whenever θ is above x . That is, $\exists x \in \Theta$ such that:

$$g(n|\theta) = 1 - g(y|\theta) = \begin{cases} 1 & \theta \leq x \\ 0 & \theta > x \end{cases}$$

Lemma 1. *For any binary signal structure $S = \{\{y, n\}, G\}$ with $\hat{\theta}(y) \geq \hat{\theta}(n)$, there exists a binary signal structure $S' = \{\{y, n\}, G'\}$ such that S' is a monotone partition, $P_{S'}(y) = P_S(y)$, and $\hat{\theta}_{S'}(y) \geq \hat{\theta}_S(y)$.*

Proof. First note that since F is continuous, there exists a monotone partition $S' = \{\{y, n\}, g'\}$ with $P_{S'}(y) = P_S(y)$, as there exists some cut-off x with $F(x) = 1 - P_S(h)$. I now show that

for any such S' , $\hat{\theta}_{S'}(h) \geq \hat{\theta}_S(h)$:

$$\begin{aligned}
P_S(y) &= P(\theta > x) \\
P_S(y|\theta \leq x)P(\theta < x) + P_S(y|\theta > x)P(\theta > x) &= P(\theta > x) \\
P_S(y|\theta \leq x)P(\theta < x) &= (1 - P_S(y|\theta > x))P(\theta > x) \\
P_S(y|\theta \leq x)P(\theta < x) &= P_S(n|\theta > x)P(\theta > x) \\
\frac{P_S(y|\theta \leq x)P(\theta < x)}{P(\theta > x)} &= P_S(n|\theta > x) \\
\frac{P_S(y|\theta \leq x)P(\theta < x)}{P_S(y)} &= P_S(n|\theta > x) \\
P_S(\theta \leq x|y) &= P_S(n|\theta > x)
\end{aligned}$$

Using this equality along with $P_S(y|\theta > x) = P_S(\theta > x|y)$ and applying the law of total expectation to the difference $\hat{\theta}_{S'}(y) - \hat{\theta}_S(y)$ gives:

$$\begin{aligned}
\hat{\theta}_{S'}(y) - \hat{\theta}_S(y) &= E(\theta|S', y) - E(\theta|S, y) \\
&= E(\theta|\theta > x) - E(\theta|S, y) \\
&= E(\theta|\theta > x, S, y)P_S(y|\theta > x) + E(\theta|\theta > x, S, n)P_S(n|\theta > x) \\
&\quad - (E(\theta|\theta > x, S, y)P_S(\theta > x|y) + E(\theta|\theta \leq x, S, y)P_S(\theta \leq x|y)) \\
&= E(\theta|\theta > x, S, n)P_S(n|\theta > x) - E(\theta|\theta \leq x, S, y)P_S(\theta \leq x|y) \\
&= P_S(\theta \leq x|y)(E(\theta|\theta > x, S, n) - E(\theta|\theta \leq x, S, y)) \\
&\geq 0
\end{aligned}$$

□

Restricting attention to binary monotone partitions reduces the planner's problem to the choice of the cut-off, x . This choice is constrained by incentive compatibility; voters must be willing to follow the recommendation they receive. That is, x must satisfy:

$$\begin{aligned}
\hat{\theta}_S(n) &= E(\theta|\theta < x) \leq v \\
\hat{\theta}_S(y) &= E(\theta|\theta > x) \geq v
\end{aligned}$$

Since $v \geq 0$, the first condition is satisfied for any $x \in \Theta$. Since $E(\theta|\theta > x)$ is increasing in x , the second condition is satisfied by any x greater than some lower bound \underline{x} :

$$\underline{x} = \min\{x \in \Theta : E(\theta|\theta > x) \geq v\}$$

Note that choosing $x = v$ is essentially equivalent to giving voters perfect information, in the sense that it leads them to vote as they would if they knew their valuation with certainty, and therefore results in an identical distribution of outcomes.

To analyze the planner's problem, it will be useful to express \bar{U} as a function of the cut-off x . All the probabilities in the expression for \bar{U} given above depend on x , as does $\hat{\theta}_S(y)$. $P_S(N)$

is the probability that $\leq \frac{Q-1}{2}$ out of Q agents vote y , and $\rho(S)$ is the probability that exactly $\frac{Q-1}{2}$ out of $Q-1$ vote y . The number of y votes follows the binomial distribution, and the probability of a y vote is $P_S(y) = 1 - F(x)$. The term $\bar{\theta}_S(y)P_S(y)$ can be written in terms of x as $\int_x^{\bar{\theta}} \theta f(\theta) d\theta$. Therefore \bar{U} can be written in terms of x as follows (using the integral form of the binomial cumulative distribution function for $P_S(N)$) :

$$\begin{aligned} \bar{U}(x) = & \left(\frac{Q+1}{2} \binom{Q}{\frac{Q-1}{2}} \int_0^{F(x)} t^{\frac{Q-1}{2}} (1-t)^{\frac{Q-1}{2}} dt \right) v \\ & + \left(\frac{Q-1}{2} \right) F(x)^{\frac{Q-1}{2}} (1-F(x))^{\frac{Q-1}{2}} \int_x^{\bar{\theta}} \theta f(\theta) d\theta \end{aligned}$$

2.2.3 Optimal Information Structures

Differentiating $\bar{U}(x)$ with respect to x and factoring out $f(x)$, binomial coefficients, and higher powers of $F(x)$ and $1 - F(x)$ gives the following first-order condition:

$$(Qv - x)F(x) + (Q-1)\left(\frac{1}{2} - F(x)\right)E(\theta|\theta > x) = 0$$

The positive term $QvF(x)$ represents the benefit of the increase in $P(N)$ when x increases, whereas the negative term (for $x > 0$) $xF(x)$ represents the change in the potential benefit of making a pivotal y vote - as x rises, the range of values of θ for which an agent will vote y falls. The term $(Q-1)\left(\frac{1}{2} - F(x)\right)E(\theta|\theta > x)$ represents the impact on expected utility of the effect a change in x has on the probability of the agent's vote being pivotal (which is maximised at $F(x) = \frac{1}{2}$).

Proposition 2. *If there exists an $x^* \in [\underline{x}, \bar{\theta}]$ that satisfies the first-order condition $\frac{d\bar{U}}{dx} = 0$, then x^* is the unique optimal choice of x for the planner. Otherwise the optimal choice is $x = \bar{\theta}$ (no information).*

Proof. As described above, $\frac{d\bar{U}}{dx} = 0$ requires:

$$(Qv - x)F(x) + (Q-1)\left(\frac{1}{2} - F(x)\right)E(\theta|\theta > x) = 0$$

When this holds, the sign of $\frac{d^2\bar{U}}{dx^2}$ is determined by the sign of the derivative of the left-hand-side:

$$(Qv - x - (Q-1)E(\theta|\theta > x))f(x) - F(x) + (Q-1)\left(\frac{1}{2} - F(x)\right)\frac{dE(\theta|\theta > x)}{dx}$$

Noting that for $x \in [\underline{x}, \bar{\theta}]$, $F(x) > 0$, the first-order condition implies that one of the following must hold:

- $Qv - x^* = 0$ and $F(x^*) = \frac{1}{2}$
- $Qv - x^* > 0$ and $F(x^*) > \frac{1}{2}$
- $Qv - x^* < 0$ and $F(x^*) < \frac{1}{2}$

In each case the derivative above is strictly negative. In the first case, it is equal to:

$$-(Q - 1)E(\theta|\theta > x^*))f(x^*) - F(x^*)$$

In the second case, all terms in the derivative are negative, since the first-order condition implies that:

$$(Qv - x^* - (Q - 1)E(\theta|\theta > x^*)) < 0$$

In the final case, note that:

$$\frac{dE(\theta|\theta > x)}{dx} = f(x) \frac{E(\theta|\theta > x) - x}{1 - F(x)}$$

Substituting this expression into the derivative above and rearranging gives:

$$(Q(v - x) - \frac{(Q - 1)}{2}(\frac{E(\theta|\theta > x) - x}{1 - F(x)}))f(x) - F(x)$$

All terms are negative for $x > v$.

It follows that if $\frac{d\bar{U}}{dx} = 0$ holds at x^* , $\frac{d^2\bar{U}}{dx^2} < 0$ also holds at x^* , and $\bar{U}(x)$ has no interior minima and at most one maximum on $[\underline{x}, 1]$.

Finally, note that:

$$\bar{U}(\bar{\theta}) = v$$

$$\bar{U}(\underline{x}) = v(P(N|\underline{x}) + \rho(\underline{x})(1 - F(\underline{x}))) < v$$

It follows that if there is no $x^* \in [\underline{x}, \bar{\theta}]$ that satisfies the first-order condition, $\bar{U}(x)$ is increasing over the interval and maximised at $x = \bar{\theta}$. \square

If the optimal choice is $x = \bar{\theta}$, the planner chooses an information structure that always results in outcome N winning! This implies that the benefits to total welfare of eliciting information about agent preferences through a vote are always outweighed by the effect of ‘tyranny of the majority’ outcomes, and the planner chooses an information structure that makes the vote redundant. The next result demonstrates that this occurs if $\bar{\theta}$ is too close to v , which means that the potential benefits of outcome Y to y voters are limited; the planner would like to ensure that only those who prefer Y by enough of a margin vote y , but if the upper bound on θ is low enough there are no such voters.

Proposition 3. *The optimal choice for the planner is $x = \bar{\theta}$ if and only if the following condition holds:*

$$v \geq \frac{Q+1}{2Q}\bar{\theta}$$

Proof.

$$\lim_{x \uparrow \bar{\theta}} (Qv - x)F(x) + (Q-1)\left(\frac{1}{2} - F(x)\right)E(\theta|\theta > x) = Q\left(v - \frac{Q+1}{2Q}\bar{\theta}\right)$$

□

While $x = \bar{\theta}$ offers the least informative solution possible, the other extreme is found at $x = v$, which leads to no ‘mistakes’ in voting. This structure gives identical outcomes to the case in which voters are perfectly informed about their valuation, so if $x = v$ is optimal, perfect information is optimal.

The clearest way in which this can be the case is when v is equal to both the mean and the median of the distribution of θ . That is, $v = 0$ and $F(v) = \frac{1}{2}$. In this case the first-order condition holds at $x = v$. Since $v = 0$, agents have no preference for N in expectation, so there is no ex ante benefit in increasing the probability that N is selected, and since $F(v) = \frac{1}{2}$ the choice of $x = v$ also maximises each agent’s probability of being pivotal. Agents therefore have no incentive to accept any chance of making a mistake in the event that they are pivotal.

This is not the only case in which perfect information may be optimal. In fact, it is a special case of a general necessary and sufficient condition, which is that the expected values of θ conditional on being greater than or less than v be symmetric.

Proposition 4. *The optimal choice for the planner is $x = v$ if and only if the following condition holds:*

$$E(\theta|\theta \geq v) - v = v - E(\theta|\theta < v)$$

That is, there exists $a \in \mathbb{R}^+$ such that:

$$E(\theta|\theta \geq v) = v + a$$

$$E(\theta|\theta < v) = v - a$$

Proof. For the first-order condition to be met at $x = v$, the following must hold:

$$vF(v) = \left(F(v) - \frac{1}{2}\right)E(\theta|\theta \geq v)$$

Rearranging for $F(v)$ gives:

$$F(v) = \frac{E(\theta|\theta \geq v)}{2(E(\theta|\theta \geq v) - v)}$$

As $E(\theta) = 0$, applying the law of total expectation gives:

$$\begin{aligned} F(v)E(\theta|\theta < v) &= -(1 - F(v))E(\theta|\theta \geq v) \\ F(v) &= \frac{E(\theta|\theta \geq v)}{E(\theta|\theta \geq v) - E(\theta|\theta < v)} \end{aligned}$$

Setting the two expressions for $F(v)$ equal and rearranging gives the condition stated in the proposition. \square

This condition for perfect information to be optimal equates the average benefit to y voters (under perfect information) of Y winning to the average benefit to n voters of N winning. Proposition 5 shows that this balance is what the optimal information structure converges to achieve as Q rises, where possible. This is only possible if $v \leq \frac{\bar{\theta}}{2}$. Otherwise once Q is sufficiently large, no information becomes optimal, per Proposition 3.

Proposition 5. *If $v \leq \frac{\bar{\theta}}{2}$, then as $Q \rightarrow \infty$, $x^* \rightarrow \hat{x}$ monotonically, where \hat{x} is the unique value in Θ such that:*

$$E(\theta|\theta \geq \hat{x}) - v = v - E(\theta|\theta < \hat{x})$$

Otherwise there exists a \hat{Q} such that for $Q \geq \hat{Q}$, $x^ = \bar{\theta}$.*

Proof. If $v = \hat{x}$, then by Proposition 4, $x^* = v = \hat{x}$ for any Q and the proposition follows trivially. Otherwise, the first-order condition can be rewritten as follows:

$$Q - 1 = \frac{x - v}{vF(x) + (\frac{1}{2} - F(x))E(\theta|\theta > x)}F(x)$$

Using the fact that $F(x) = \frac{E(\theta|\theta > x)}{E(\theta|\theta > x) - E(\theta|\theta < x)}$, this can be further rewritten as:

$$Q - 1 = \frac{2(x - v)}{(v - E(\theta|\theta < x)) - (E(\theta|\theta > x) - v)}$$

The denominator of the right-hand side is clearly decreasing in x and equal to 0 at $x = \hat{x}$. Since $Q - 1$ is positive, this allows two possibilities for x^* :

- $x^* > v$ and $x^* < \hat{x}$.
- $x^* < v$ and $x^* > \hat{x}$.

It follows that if $v < \hat{x}$, then $v < x^* < \hat{x}$ for all Q . Alternatively, if $v > \hat{x}$, then $\hat{x} < x^* < v$ for all Q .

In the first case, the right-hand side fraction is increasing in x , while in the second case it is decreasing. It follows that, in the first case, for any $x' \in (v, \hat{x})$ there exists a \bar{Q} such that:

$$\bar{Q} - 1 > \frac{2(x' - v)}{(v - E(\theta|\theta < x')) - (E(\theta|\theta > x') - v)}$$

Therefore for all $Q > \bar{Q}$, $\hat{x} > x^* > x'$, and it follows that x^* converges to \hat{x} as Q goes to infinity. A symmetric argument applies in the second case. \square

2.2.4 Alternative majority thresholds

The results of the previous section generalise to votes that use a threshold other than simple majority rule. Consider a voting rule under which outcome Y is implemented if and only if the number of y votes is greater than $k \leq Q$. The simple majority is given by $k = \frac{Q-1}{2}$. The optimality of binary monotone partitions is unchanged in this case, and the expected utility of a voter in terms of the partition cut-off x becomes:

$$\begin{aligned} \bar{U}(x) = & \left((Q-k) \binom{Q}{k} \int_0^{F(x)} t^{Q-k-1} (1-t)^k dt \right) v \\ & + \binom{Q-1}{k} F(x)^{Q-k-1} (1-F(x))^k \int_x^1 \theta f(\theta) d\theta \end{aligned}$$

As before, a simplified first-order condition for an optimal internal solution can be derived:

$$(Qv - x)F(x) + (Q-1) \left(\frac{Q-k-1}{Q-1} - F(x) \right) E(\theta|\theta > x) = 0$$

Corresponding results to those of the previous section follow. In particular, perfect information is optimal if and only if:

$$k(E(\theta|\theta > v) - v) = (Q-k-1)(v - E(\theta|\theta < v))$$

This is simply a weighted version of the condition in the simple majority case, in which $k = Q - k - 1$. It equalises the expected impacts of implementing outcome Y over N on yes voters and no voters weighted by the proportion of each type in the population when a deciding vote is cast.

In considering the limit information structure as Q becomes large, it is appropriate to treat the rule as requiring y votes greater than a fixed fraction cQ , $c \in (0, 1)$. Then for each Q , k is the largest integer that does not exceed cQ . Proposition 6 demonstrates that the convergence result for simple majority rule generalises; the optimal threshold x converges to a point at which the expected benefits to each type of voter are balanced when weighted according to the proportions of each type present in the population when the vote is exactly balanced. As in the simple majority case, when perfect information is not optimal because it does not appropriately balance the expected impacts, then as Q becomes large, x^* approaches the point which does achieve that balance (if this point exists; as before, it depends on v not being too close to $\bar{\theta}$).

Proposition 6. *If for each Q the voting threshold k is the largest integer less than cQ , $c \in (0, 1)$, and $v \leq c\bar{\theta}$, then as $Q \rightarrow \infty$, $x^* \rightarrow \hat{x}_c$ monotonically, where \hat{x}_c is the unique value in Θ such that:*

$$c(E(\theta|\theta > \hat{x}_c) - v) = (1-c)(v - E(\theta|\theta < \hat{x}_c))$$

If $v > c\bar{\theta}$, then there exists a \hat{Q} such that for $Q \geq \hat{Q}$, $x^ = \bar{\theta}$.*

The proof is analogous to the proof of Proposition 5.

2.3 Abstention

The results of the previous section show that with mandatory voting, the inefficiencies of the voting system can be somewhat mitigated by restricting voter information if and only if the distribution of valuations is asymmetric in the sense given by Proposition 4. However, inefficiencies are still present even if this asymmetry is not (that is, if $x^* = v$). The proposition only implies that in such cases, restricting voter information cannot help. The benefits of restricting information with mandatory voting come from causing either y or n voters (crucially, *not both*) who are relatively close to being indifferent to vote ‘incorrectly’. If the expected impacts of the policy on y and n voters are already exactly balanced when those voters are perfectly informed, this does not help, since the benefits in terms of increasing the probability that one outcome happens ‘when it should’ (i.e. when it maximises total utility given the true valuations) are exactly cancelled out by the losses from the increased probability of that outcome happening when it should not.

If the assumption of mandatory voting is relaxed, then it is possible to have relatively-indifferent voters abstain from voting, rather than relying on them voting incorrectly. In this case, the planner may benefit by restricting information even when the distribution of valuations is symmetric. To demonstrate this, in this section I make the following modifications to the model:

- Voters’ action set is $\{y, n, b\}$.
- The vote is conducted by comparing the numbers of y and n votes. b votes are ignored.
- Voters who are indifferent between outcomes choose b .
- $v = 0$.
- θ is distributed symmetrically around 0; that is, $\underline{\theta} = -\bar{\theta}$ and $f(\theta) = f(-\theta)$.
- Ties are now possible and are broken by 50-50 randomization.

The model is otherwise identical to the previous section. Assuming symmetry simplifies the discussion while focusing on a case in which perfect information is optimal under mandatory voting. Note that whether or not symmetry is assumed, the introduction of abstentions makes the planner more powerful; it does not significantly restrict what she can achieve compared to mandatory voting. Binary monotone partitions with a cut-off in $(\underline{x}, \bar{\theta}]$ result in the same outcomes as before, since their messages never leave voters indifferent. The effect of a binary monotone partition with cut-off \underline{x} does change, but the proof of Proposition 2 demonstrates that this information structure is never optimal for the planner under mandatory voting.

I denote the *ex ante* expected utility of an agent given this modified setting and a signal structure S by $\bar{U}_a(S)$. Note that since incentive compatibility requires that $E(\theta|a = b) = 0$, $\bar{U}_a(S)$ can be expressed similarly to $\bar{U}(S)$:

$$\bar{U}_a(S) = \rho(S)P_S(y)\hat{\theta}_S(y)$$

Since there are now three possible actions, the planner may now choose to use three-message information structures, with the caveat that one message must induce abstention ($\hat{\theta}_S = 0$):

Proposition 7. *Let $S = \{M, g\}$ be a signal structure, with $|M| > 3$. Then there exists a signal structure $S' = \{M', g'\}$ such that $M' = \{y, n, b\}$, $\bar{U}_a(S') = \bar{U}_a(S)$*

Proof. The proof follows the same steps as that of Proposition 1, except that messages in M that result in indifference (i.e. $\hat{\theta}_S(m) = 0$) result in abstention (action b) rather than randomization and are mapped to a distinct b message in M' , with $\hat{\theta}_{S'}(b) = 0$. \square

Define a **three-message symmetric monotone partition** as a signal structure $S = (M, g)$ with $M = \{y, n, b\}$, for which there exists a $\gamma \in [0, 1]$ such that:

$$\begin{aligned} g(n|\theta) &= \begin{cases} 1 & \theta \leq -x \\ 0 & \text{otherwise} \end{cases} \\ g(b|\theta) &= \begin{cases} 1 & -x < \theta < x \\ 0 & \text{otherwise} \end{cases} \\ g(y|\theta) &= \begin{cases} 1 & \theta \geq x \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Note that given $E(\theta) = 0$ and the symmetry of the distribution of θ , this implies that $\hat{\theta}_S(b) = 0$, that $\hat{\theta}(n) = -\hat{\theta}(y)$, and that:

$$P_S(y) = P_S(n) = \frac{1 - P_S(b)}{2}$$

The next set of results show that attention can be restricted to three-message symmetric monotone partitions, as some such signal structure is always weakly optimal among three-message structures.

Lemma 2. *For a fixed $P_S(b)$, the probability that a given y vote is pivotal, $\rho(S)$, is maximised at $P_S(y) = P_S(n)$.*

Proof. $\rho(S)$ can be written as follows:

$$\rho(S) = \sum_{k=0}^{Q-1} T_S(k)$$

where $T_S(k)$ gives the probability that the voter is pivotal and that exactly k of the other voters abstain. For $Q - k$ odd:

$$T_S(k) = \frac{(Q - 1 - k)!}{k! \left(\frac{Q-1-k}{2}\right)!^2} P_S(b)^k P_S(y)^{\frac{Q-1-k}{2}} (1 - P_S(b) - P_S(y))^{\frac{Q-1-k}{2}}$$

For $Q - k$ even⁴:

$$T_S(k) = \frac{(Q - 1 - k)!}{k! \left(\frac{Q-k}{2}\right)! \left(\frac{Q-k}{2} - 1\right)!} P_S(b)^k P_S(y)^{\frac{Q-k}{2}-1} (1 - P_S(b) - P_S(y))^{\frac{Q-k}{2}-1}$$

In either case, for a fixed $P_S(b)$, $T_S(k)$ is maximised for each k when $P_S(y) = P_S(n) = \frac{1-P_S(b)}{2}$. \square

Lemma 2 demonstrates the intuitive conclusion that the probability of a given vote being pivotal is maximised when the probability of y and n votes are equal, as this maximises the probability of a tie or near-tie among the other non-abstaining voters.

The next result shows that signal structures can be improved upon by ‘unmixing’ the two vote-inducing messages, while leaving the abstention-inducing message untouched.

Lemma 3. *For any signal structure $S = \{\{y, n, b\}, G\}$ with:*

$$\hat{\theta}_S(y) \geq \hat{\theta}_S(b) = 0 \geq \hat{\theta}_S(n)$$

there exists a signal structure $S' = \{\{y, n, b\}, G'\}$ and a cut-off $x \in \Theta$ such that:

- $P_{S'}(y) = P_S(y)$, $P_{S'}(b) = P_S(b)$, and $P_{S'}(n) = P_S(n)$
- $\hat{\theta}_{S'}(y) \geq \hat{\theta}_{S'}(b) = 0 \geq \hat{\theta}_{S'}(n)$
- $P_{S'}(y|\theta < x) = P_{S'}(n|\theta \geq x) = 0$
- $\bar{U}_a(S') \geq \bar{U}_a(S)$

Proof. Begin by choosing x to solve the following equation:

$$P_S(y) = P_S(\theta \geq x, m \neq b)$$

A solution exists in Θ as the right hand side is continuous, weakly decreasing in x , equal to $P_S(h) + P_S(l)$ at $x = \underline{\theta}$, and equal to 0 at $x = \bar{\theta}$. Now construct g' as follows:

$$g'(b|\theta) = g(b|\theta)$$

⁴To understand the derivation of this expression, note that the probability of making a pivotal y vote is $P_S(Y|y) - P_S(Y|n)$. When the number of other voters is odd, switching from n to y can impact $P(Y)$ by causing a tie (N ahead by one vote among the other voters) or by breaking one (Y ahead by one vote among the other voters). The increase in $P(Y)$ is $\frac{1}{2}$ in both cases. This leads to the expression given.

$$g'(n|\theta) = \begin{cases} g(n|\theta) + g(y|\theta) & \theta < x \\ 0 & \text{otherwise} \end{cases}$$

$$g'(y|\theta) = \begin{cases} 0 & \theta < x \\ g(n) + g(y) & \text{otherwise} \end{cases}$$

It is straightforward to check that g' fulfills the conditions on $P_{S'}(y)$ and $P_{S'}(n)$ given in the statement of the proposition. Since the probabilities of each action under S are preserved by S' , $\rho(S') = \rho(S)$. It follows that to prove $\bar{U}_a(S') \geq \bar{U}_a(S)$, it is sufficient to show that $\hat{\theta}_{S'}(y) \geq \hat{\theta}_S(y)$. Note that $\hat{\theta}_{S'}(y) = E(\theta|S, \theta \geq x, m \neq z)$. Given this, $\hat{\theta}_{S'}(y) \geq \hat{\theta}_S(y)$ can be shown using the same method as in the proof of Lemma 1. \square

Once the vote-inducing signals have been ‘unmixed’ in this way, the next result shows that when abstentions occur, it should be the most ambivalent voters who abstain. That is, the abstention-inducing message should be sent if and only if valuations fall into an interval that is symmetrical around 0.

Lemma 4. *For any signal structure $S = \{\{y, n, b\}, G\}$ and cut-off x with:*

$$\hat{\theta}_S(y) \geq \hat{\theta}_S(b) = 0 \geq \hat{\theta}_S(n)$$

and:

$$P_S(y|\theta < x) = P_S(n|\theta \geq x) = 0$$

there exists a signal structure $S' = \{\{y, n, b\}, G'\}$ and cut-off $\gamma \geq 0$ such that:

- $\hat{\theta}_{S'}(y) \geq \hat{\theta}_{S'}(b) = 0 \geq \hat{\theta}_{S'}(n)$
- $P_{S'}(y|\theta < x) = P_{S'}(n|\theta \geq x) = 0$
- $P_{S'}(b) = P_S(b)$
- $P_{S'}(b|\theta \in [-\gamma, \gamma]) = 1$ and $P_{S'}(b|\theta \notin [-\gamma, \gamma]) = 0$
- $\hat{\theta}_{S'}(y)P_{S'}(y) \geq \hat{\theta}_S(y)P_S(y)$

Proof. Begin by choosing γ to solve:

$$F(\gamma) - F(-\gamma) = P_S(b)$$

Now construct g' as follows:

$$g'(b|\theta) = \begin{cases} 1 & \theta \in [-\gamma, \gamma] \\ 0 & \text{otherwise} \end{cases}$$

$$g'(n|\theta) = \begin{cases} 1 & \theta < x, \theta \notin [-\gamma, \gamma] \\ 0 & \text{otherwise} \end{cases}$$

$$g(y|\theta) = \begin{cases} 1 & \theta > x, \theta \notin [-\gamma, \gamma] \\ 0 & \text{otherwise} \end{cases}$$

To see that $\hat{\theta}_{S'}(y)P_{S'}(y) \geq \hat{\theta}_S(y)P_S(y)$, note that:

$$\begin{aligned} \hat{\theta}_{S'}(y)P_{S'}(y) - \hat{\theta}_S(y)P_S(y) &= (1 - F(x)) \left(E(\theta|S, \theta > x, b)P_S(b|\theta > x) \right. \\ &\quad \left. - E(\theta|S', \theta > x, b)P_{S'}(b|\theta > x) \right) \end{aligned}$$

Assume that:

$$E(\theta|S', \theta > x, b)P_{S'}(b|\theta > x) > E(\theta|S, \theta > x, b)P_S(b|\theta > x)$$

This implies $\gamma > x$, since otherwise $P_{S'}(b|\theta > x) = 0$. Given that, note that among all subsets Z of $[x, \bar{\theta}]$ with $P(\theta \in Z|\theta > x) = P_{S'}(b|\theta > x)$, the interval $[x, \gamma]$ minimizes $E(\theta|\theta \in Z)$, which implies that $E(\theta|S', \theta > x, b) \leq E(\theta|S, \theta > x, b)$. Therefore, for the assumption to hold it must be the case that:

$$P_{S'}(b|\theta > x) > P_S(b|\theta > x)$$

Applying the law of total expectation to the original assumption gives:

$$E(\theta|S, \theta > x, b)P_S(b|\theta < x) > E(\theta|S', \theta < x, b)P_{S'}(b|\theta < x)$$

However, it is also the case that among all subsets Z of $[\underline{\theta}, x]$ with $P(\theta \in Z|\theta > x) = P_{S'}(b|\theta < x)$, $[-\gamma, x]$ maximizes $E(\theta|\theta \in Z)$, which (given that the expectations in the inequality above are negative) implies that:

$$P_{S'}(b|\theta < x) > P_S(b|\theta < x)$$

This implies that $P_{S'}(b) > P_S(b)$, which is a contradiction. It follows that

$$E(\theta|S', \theta > x, b)P_{S'}(b|\theta > x) \leq E(\theta|S, \theta > x, b)P_S(b|\theta > x)$$

and therefore $\hat{\theta}_{S'}(y)P_{S'}(y) \geq \hat{\theta}_S(y)P_S(y)$. □

Finally, combining the preceding results shows that symmetric monotone partitions are weakly optimal.

Proposition 8. *For any signal structure $S = \{\{y, n, b\}, G\}$ with:*

$$\hat{\theta}_S(y) \geq \hat{\theta}_S(b) = 0 \geq \hat{\theta}_S(n)$$

there exists a signal structure $S' = \{\{y, n, b\}, G'\}$ such that S' is a three-message symmetric monotone partition and $\bar{U}_a(S') \geq \bar{U}_a(S)$.

Proof. Begin by constructing an intermediate signal structure $S'' = \{\{y, n, b\}, g''\}$ using the steps given in the proofs of Lemma 3 and 4, resulting in a structure with cut-offs x and γ such that b is sent if $-\gamma \leq \theta \leq \gamma$, n is sent if $\theta < \min\{x, -\gamma\}$, and y is sent if $\theta > \max\{x, \gamma\}$. Per the preceding proofs, this construction also gives:

- $P_{S''}(b) = P_S(b)$
- $\theta(\hat{y})_{S''}P_{S''}(y) \geq \theta(\hat{y})_SP_S(y)$

Now construct g' as follows:

$$g'(b|\theta) = \begin{cases} 1 & \theta \in [-\gamma, \gamma] \\ 0 & \text{otherwise} \end{cases}$$

$$g'(n|\theta) = \begin{cases} 1 & \theta < -\gamma \\ 0 & \text{otherwise} \end{cases}$$

$$g(y|\theta) = \begin{cases} 1 & \theta > \gamma \\ 0 & \text{otherwise} \end{cases}$$

This construction gives $P_S(y) = P_S(n)$. By Lemma 2, it follows that the $\rho(S') \geq \rho(S)$ (since both S and S' induce the same abstention probability, $P_{S'}(b) = P_S(b)$). To see that $\hat{\theta}_{S'}(y)P_{S'}(y) \geq \hat{\theta}_S(y)P_S(y)$, note that if $x < -\gamma$:

$$\hat{\theta}_{S'}(y)P_{S'}(y) - \hat{\theta}_S(y)P_S(y) = -E(\theta|x < \theta \leq -\gamma)P(x < \theta \leq -\gamma)$$

If $x > \gamma$, then:

$$\hat{\theta}_{S'}(y)P_{S'}(y) - \hat{\theta}_S(y)P_S(y) = E(\theta|\gamma \leq \theta < x)P(\gamma \leq \theta < x)$$

Finally, if $x \in [-\gamma, \gamma]$ then:

$$\hat{\theta}_{S'}(y)P_{S'}(y) = \hat{\theta}_S(y)P(y)$$

□

Restricting attention to symmetric monotone partitions reduces the problem, once again, to the choice of a single variable, γ . If S is a symmetric monotone partition, $\bar{U}_a(S)$ can be written as $\bar{U}_a(\gamma)$. This leads to the central result of this section; that with symmetry and abstentions, for a sufficiently large number of voters, giving voters perfect information is *never* optimal.

Proposition 9. $\bar{U}_a(\gamma)$ is not maximised on $[0, \bar{\theta}]$ at $\gamma = 0$ for sufficiently large Q .

Proof. Note that:

$$U_a(\gamma) = \rho(\gamma) \int_{\gamma}^{\bar{\theta}\theta f(\theta)} d\theta$$

$$\frac{dU_a}{d\gamma} = \frac{d\rho}{d\gamma} \int_{\gamma}^{\bar{\theta}\theta f(\theta)} d\theta - \rho(\gamma) (\gamma f(\gamma))$$

Clearly the second term is 0 at $\gamma = 0$, so it is sufficient to show that $\frac{d\rho}{d\gamma}$ is positive at $\gamma = 0$. Define a function $h(Q, k)$ for integers Q and k as follows:

$$h(n, k) = \begin{cases} \frac{(n-1)!}{k! \left(\frac{n-1-k}{2}\right)! \left(\frac{n-1+k}{2}\right)!} & (n-k) \text{ odd} \\ \frac{(n-1)!}{k! \left(\frac{n-k}{2}\right)! \left(\frac{n-2-k}{2}\right)!} & (n-k) \text{ even} \end{cases}$$

Then $\rho(\gamma)$ can be written as follows:

$$\rho(\gamma) = \sum_{k=0}^{Q-1} h(Q, k) (1 - F(\gamma))^{Q-1-k} (2F(\gamma) - 1)^k$$

Each term in the summation gives the probability that k of the other voters abstain, and that the vote among the remaining $n - k - 1$ voters is tied (for $n - k - 1$ even) or leaves one outcome ahead by a single vote (for $n - k - 1$ odd). It follows that:

$$\begin{aligned} \frac{d\rho}{d\gamma} &= \sum_{k=0}^{Q-1} h(Q, k) f(\gamma) \left[2k(1 - F(\gamma))^{Q-1-k} (2F(\gamma) - 1)^{k-1} \right. \\ &\quad \left. - (Q - 1 - k)(1 - F(\gamma))^{Q-2-k} (2F(\gamma) - 1)^k \right] \end{aligned}$$

At $\gamma = 0$, $F(\gamma) = \frac{1}{2}$ and all but two terms are 0. For even Q , $h(Q, 1) = \frac{Q}{2}h(Q, 0)$, which, assuming $Q > 2$, gives:

$$\begin{aligned} \frac{d\rho(0)}{d\gamma} &= 2h(Q, 1)f(0)\left(\frac{1}{2}\right)^{Q-2} - (Q - 1)h(Q, 0)\left(\frac{1}{2}\right)^{Q-2} \\ &= h(Q, 0)f(0)\left(\frac{1}{2}\right)^{Q-2} \\ &> 0 \end{aligned}$$

For odd Q , $h(Q, 1) = \frac{Q-1}{2}h(Q, 0)$, which gives $\frac{d\rho(0)}{d\gamma} = 0$, and consequently $\frac{dU_a(0)}{d\gamma} = 0$. Checking the second derivative demonstrates that this is a local minimum if Q is sufficiently large:

$$\begin{aligned} \frac{d^2U_a}{d\gamma^2} &= \frac{d^2\rho}{d\gamma^2} \int_{\gamma}^{\bar{\theta}} \theta f(\theta) d\theta - \frac{d\rho}{d\gamma} (\gamma f(\gamma)) - \rho(\gamma) (\gamma f'(\gamma) + f(\gamma)) \\ \frac{d^2U_a(0)}{d\gamma^2} &= \frac{d^2\rho(0)}{d\gamma^2} \int_0^{\bar{\theta}} \theta f(\theta) d\theta - \rho(0) f(0) \\ &= (Q - 1)f(0)^2 \left(\frac{1}{2}\right)^{Q-3} h(Q, 0) \int_0^{\bar{\theta}} \theta f(\theta) d\theta - h(Q, 0) \left(\frac{1}{2}\right)^{Q-1} f(0) \\ &= f(0) \left(\frac{1}{2}\right)^{Q-3} ((Q - 1)E(\theta|\theta > 0)f(0) \int_0^{\bar{\theta}} \theta f(\theta) d\theta - \frac{1}{4}) \end{aligned}$$

□

2.4 Conclusion

The model developed in this paper demonstrates that a utilitarian social planner may not want voters to be perfectly informed about how the possible outcomes of a vote will affect them. Of course, in reality, it will often be more of a concern that voters are too poorly informed. The results given here simply sound a note of caution about the problems (from a utilitarian perspective) with the voting mechanism itself, and how information that is excessively precise may exacerbate those problems by allowing those who are minimally impacted by the voting outcome to have a disproportionate effect on it (and, with abstention, by encouraging them to vote in the first place!).

I have largely abstracted away from the possibility of common values among the voters in this model, though it is possible to interpret it as taking place after a common value element has been ‘parceled out’ and publicly signaled, determining the value of the parameter v . When the effect of information on voter valuations is likely to be highly correlated across voters, it will typically be beneficial to signal that information accurately. As described in the introduction, this model pertains to information that may result in a wide range of valuations for any given voter, depending on their preferences. It is worth noting that the ‘voters’ in this model can be interpreted as equally-sized groups of individual voters, each group having its own shared set of preferences. A possible extension of the model along these lines would add weights to each voter, in both the planner’s objective and the voting mechanism itself.

Another direction for future work is to consider non-binary votes. Once there are more than two outcomes, strategic voting enters the picture. The possible effects of the information structure on voters’ incentives to vote strategically, and the resulting impacts on welfare, make for a promising line of inquiry.

Bibliography

- [1] Alonso, Ricardo, and Odilon Câmara. (2016) “Persuading voters.” *American Economic Review* 106.11: 3590-3605
- [2] Azrieli, Yaron, and Semin Kim, S., 2014. “Pareto Efficiency and Weighted Majority Rules.” *International Economic Review*, 55(4), pp.1067–1088.
- [3] Bardhi, Arjada, and Yingni Guo. (2018) “Modes of persuasion toward unanimous consent.” *Theoretical Economics* 13.3: 1111-1149
- [4] Borgers, Tilman. (2004) “Costly voting.” *American Economic Review* 94.1: 57-66
- [5] Buchanan, James M., and Gordon Tullock. (1962) *The Calculus of Consent*, Ann Arbor: University of Michigan Press
- [6] Casella, Alessandra. (2005) “Storable Votes.” *Games and Economic Behavior* 51.2: 391-419.
- [7] Casella, Alessandra, and Andrew Gelman. (2008) “A simple scheme to improve the efficiency of referenda.” *Journal of Public Economics* 92.10-11: 2240-2261.
- [8] Casella, Alessandra, and Thomas Palfrey. (2019) “Trading Votes for Votes. A Dynamic Theory.” *Econometrica* 87.2: 631-652.
- [9] Kamenica, Emir, and Matthew Gentzkow. (2011) “Bayesian persuasion.” *The American Economic Review* 101, no. 6: 2590-2615.
- [10] Lalley, Steven P., and E. Glen Weyl. (2018) “Quadratic voting: How mechanism design can radicalize democracy.” *AEA Papers and Proceedings* Vol. 108
- [11] Van der Straeten, Karine and Takuro Yamashita (2019) “On the veil-of-ignorance principle: welfare-optimal information disclosure in voting.” *Mimeo*

Chapter 3

Information Sharing in Continuous Two-Player Games

3.1 Introduction

Agents engaged in strategic interactions will often have preferences over what their opponents believe about them, since this will influence equilibrium play. One way in which agents can affect those beliefs is to reveal, hide, or partially reveal private information to which they have access. In this paper, I consider two-player games in which each player can select the information structure through which their opponent learns about the player's own payoff type, but has to commit to this choice *before* learning their own payoffs. One can think of this as the agent making a choice about how 'public' the investigative process through which they learn their own payoffs is.

This structure has a precedent in the extensive literature on information sharing in oligopoly, which goes back to Novshek and Sonnenschein (1982). In these models, firms engaged in oligopolistic competition make a commitment ahead of time to reveal or conceal information about market demand or private costs (the decision is typically considered to be implemented through trade associations). Whether firms reveal or conceal information depends on the information in question and the structure of the underlying game (Cournot versus Bertrand competition, complements versus substitutes). I review this and other related literature in Section 3.1.1 below.

The model I present allows for a greater range of information structures than in the information sharing literature, which assumes either a binary choice between revelation and concealment, or allows for the addition of unbiased noise (by having the agents choose the variance of an error term in a signal, or choose the proportion of private signals to be added to a public pool). I consider the possibility that agents can engage in 'persuasion' of their opponent, in the sense of Kamenica and Gentzkow (2011). In line with that paper and the in-

formation design literature that follows it, I allow agents to choose a stochastic mapping from their possible payoff types to messages that may be sent to their opponents. To see how this allows for possibilities not encompassed by adding noise, consider a Cournot duopolist who may have high or low costs. Considering ‘persuasive’ information structures allows for the duopolist adopting asymmetric information structures, with signaling behavior that varies dependent on the realised costs; for example, a structure that sends a ‘low’ message whenever the firm realises low costs, but mixes between ‘low’ and ‘high’ messages whenever it realises high costs.

The model consists of a two-stage simultaneous-move game with two players. Payoffs are initially uncertain, with each player’s payoff type to be drawn from a set of two possibilities according to a common prior. In the first stage, each player simultaneously chooses and commits to an information structure. Players then observe the other player’s first-stage choice, learn their own payoff type with certainty, and receive a signal about the payoff type of their opponent *generated by the information structure their opponent chose in the first stage*. In the second stage, the players engage in a simultaneous game in which the action space is an interval, with payoffs determined by the second-stage actions and the payoff types. So in the Cournot example, each firm learns their own cost with certainty before choosing output (but *after* choosing an information structure), and what they learn about their opponent’s cost depends on the information structure that their opponent chose in the first stage.

In the first stage of this game, the players act as information designers, but interacting ones. The information design problem each player faces in determining their first-stage best response is shaped by the strategy of their opponent, because the players care about the second-stage interaction between the receivers they are each transmitting information to (their respective opponent’s second-stage selves). The model is therefore a special case of interactive information design as described in Koessler, Laclau and Tomala (2020), which analyzes the general case of interacting designers who care about the outcome of a game played among a number of receivers. That paper is primarily concerned with the existence of equilibrium. As the authors point out, single-sender information design tends to use a sender’s-preferred-outcome tie-breaking rule to assure the existence of a solution to the sender’s problem. This does not transfer to the case of multiple senders, and the possibility of sender expected payoffs that are discontinuous in the posterior beliefs of receivers raises a question as to equilibrium existence in the first stage. Koessler et al. give an existence proof that relies on the presence of a correlating device, essentially requiring that the players be allowed to play correlated equilibria in the second stage.

For the model of this paper, I provide sufficient conditions on payoffs to guarantee existence without a correlating device. These conditions are a stronger version of standard conditions that guarantee equilibrium uniqueness in the second stage for any given posterior beliefs, and they encompass a large class of concave games, including typical quadratic forms of Cournot and Bertrand oligopoly games. They additionally guarantee the existence of *Markovian* equilibria; I adopt this terminology from Koessler et al. (2020) to describe equilibria in which players’ second-stage strategies are conditioned only on their posterior beliefs about

the state of the world (in this case, their knowledge of their own payoff type and their beliefs about their opponent's type) and do not otherwise depend on the first-stage choices.

Markovian equilibria are of particular interest because they allow analysis of a player's first-stage problem using the tools of the information design literature. Given Markovian second-stage strategies and a first-stage strategy for one player, the second player's expected payoff can be written as a function of the first player's posterior belief, which I refer to as the player's *ex-ante value function*. This gives the expected payoff of the second player *conditional on the first player receiving information that induces a given posterior belief*. The concave closure of the ex-ante value function then determines the second player's best response, per Kamenica and Gentzkow (2011), and analysis of the function and its convexity properties can reveal whether particular forms of payoff uncertainty and the resulting second-stage games lead to revelation, concealment, or other behavior in the first stage, and why.

I demonstrate that the derivatives of this first-stage value function can be split into two distinct effects, which represent two different impacts on the conditional expected payoff of a player of a shift in their opponent's posterior belief. One effect arises straightforwardly from the change in expected second-stage play; any change in the beliefs of either player typically leads to changes in the equilibrium second-stage actions of both players. I call this the *equilibrium-action effect*. The other effect is due to the *ex ante* implications for a player's own expected preferences of a shift in the opponent's posterior. Since the players observe the information structures chosen by their opponent and update their beliefs rationally, an increase in an opponent's posterior belief that a player has a particular payoff type implies an increase in the probability that the player does have that payoff type, conditional on the opponent's posterior. I call this the *own-type effect*.

Returning to the Cournot example will help to clarify the role of these effects. An increase in one player's posterior belief that the other has a low cost of production leads to reduced equilibrium production for the former player and corresponding increased equilibrium production (and an increased expected payoff) for the latter. This is the expected-action effect. The only way for the latter player to bring about such a change in the posterior that will result from a given message is to increase the probability that the message is sent when they do have low costs, or to decrease the probability that it is sent when they in fact have high costs - effectively, to increase the probability that they have low costs conditional on that message being sent. Because they produce more when their costs are low, their expected payoff conditional on the message being sent consequently rises. This is the own-type effect.

To demonstrate how this approach can shed light on particular applications, I analyze a variety of games with quadratic payoffs, which allow for explicit derivation of the equilibrium second-stage actions. For a certain class of games which encompasses the case of Cournot oligopoly with linear cost uncertainty as described above (for both substitute and complementary goods), I show that both effects are always positive. This guarantees a convex ex-ante value function, and consequently full revelation of types as a dominant-strategy equilibrium.

This result is in line with the information sharing literature, and shows that the finding of full revelation of cost information under Cournot competition and linear marginal costs is

robust to allowing for persuasive information structures. Additionally, it clarifies the structural features underlying those results. The expected benefit a firm in Cournot competition derives from an increase in an opponent's belief that the firm has low costs is increasing in that belief, because it is proportional to the firm's own expected production level, and that production level is decreasing in the firm's expected costs (own-type effect) and in the opponent's expected production level (equilibrium-action effect). Since the expected benefit is increasing, it is optimal for the firm to 'gamble' by choosing an information structure that mixes between extreme posterior beliefs for its opponent - that is, a revealing information structure.

The approach I use also clarifies why contrasting results of full concealment have been found for differentiated Bertrand competition. I show that the key structural difference that leads to concealment of costs in Bertrand competition versus revelation under Cournot is the presence of an 'externality' in Bertrand competition; that is, a component of each player's payoff that is determined by their opponent's action independently of the player's own action. This externality is the impact on a player's total costs of production of their opponent's choice of price (through its effect on demand), holding the player's own price fixed. This varies with marginal cost and enters into the own-type effect, changing its sign and reversing optimal first-stage behavior. In Bertrand competition with substitute goods, for example, a player benefits from an increase in their opponent's belief that the player has high costs, as this leads the opponent to expect a higher price from the first player and to raise their own price accordingly. However, this benefit is decreasing in the probability that the player actually has high costs - there is an increasing component as their own equilibrium price rises, but this is outweighed by the falling benefit of increased demand due to rising expected marginal cost.

In addition to providing insight into these well-treated cases, I consider several further classes of games that demonstrate how first-stage behavior in other settings may not be as straightforward. In Cournot competition with quadratic costs, when the uncertainty is only over the scale of the cost function (i.e. there is one type draw for which the linear and quadratic coefficients of the cost function are both larger than for the other possible type), the linear result generalises and full revelation still obtains. However, I show that if the uncertainty is between cost functions in which one has a larger linear coefficient and the other a larger quadratic component (i.e. the marginal cost functions cross), then cost information may be concealed rather than revealed, as a result of equilibrium production choices that vary nonlinearly in beliefs.

I also consider alternative forms of uncertainty that serve to demonstrate that the dominant-strategy, fully-revealing or fully-concealing outcomes of the settings described above should not be expected in general. I consider a class of games for which the initial uncertainty is over the cross-effect - for each player, there is a possibility that the opponent's action is a strategic complement and an alternative possibility that it is a strategic substitute. I show that for some range of priors this leads to fully-revealing behaviour, and for another to fully-concealing behaviour, but there is also an intermediate range for which multiple equilibria exist. When the prior falls in this latter range, revelation is the best response

to revelation, but concealment is the best response to concealment, and there also exists a third symmetric equilibrium in which both players choose a partially-revealing information structure. This demonstrates that unique and continuously-related equilibria in the second-stage games do not guarantee uniqueness in the overall game, and that ‘persuasive’ play is possible in equilibrium.

Section **3.2** introduces the model by working through a basic example: Cournot duopoly with homogeneous goods and uncertainty over linear costs. Section **3.3** develops the general model, including a set of sufficient conditions on payoffs for equilibrium existence, and the general form of the ex-ante value function and its second derivative (broken down into the expected-action and own-type effects). Section **3.4** analyzes various forms of uncertainty for games with quadratic payoffs, identifying classes of games with fully-revealing and fully-concealing outcomes which encompass the oligopoly games with cost uncertainty considered in the information sharing literature, and considering games with cross-effect uncertainty for which the outcomes are less straightforward and may include partially-revealing information structures.

3.1.1 Related Literature

This paper contributes to the literature on information design. Kamenica and Gentzkow (2011) give a general analysis of information design or persuasion in the single-sender, single-receiver case, using the ‘concavification’ approach I adopt in this paper (which was also used in the context of repeated games in Aumann, Maschler and Stearns (1995) and the related literature). Bergemann and Morris (2016), Taneva (2019), Mathevet, Perogo and Taneva (2020) and others extend the analysis to settings with multiple, interacting receivers¹. Bergemann and Morris (2019) and Kamenica (2019) review many of the varying settings considered in the literature.

The literature on information design settings with multiple senders is less extensive. Gentzkow and Kamenica (2016, 2017) and Li and Norman (2018) consider multiple senders with access to the same information about a singular state of the world face a single receiver, and focus on the conditions under which moving from single to multiple senders makes the receiver more informed. Albrecht (2017), Au and Kawai (2019) and Boleslavsky and Cotton (2015) all analyze particular settings in which multiple senders with information about different things (i.e. different components of the overall state of the world, as in this paper), all with a single receiver. Koessler, Laclau, Renault and Tomala (2019) consider two senders and a single receiver in the context of a multistage game. As described above, Koessler et al. (2020) consider a general multi-sender, multi-receiver game, of which the model of this paper is a particular case.

Two papers connect this literature to models of information sharing in oligopoly games, which

¹Although the model of this paper includes multiple receivers, the information design problem is essentially a single-receiver one, since each sender is choosing the information structure for a single receiver (their opponent).

provide many of the examples I consider here. Bergemann and Morris (2013) show that the result of Clarke (1983) that firms will not share demand information in Cournot competition with homogeneous goods is not robust to allowing for a full range of possible information structures. Their work is distinct from this paper both in that they consider ‘public value’ demand information, whereas I focus on private values (typically cost information in an oligopoly interpretation), and in that they use information structures that go beyond the scope of what I allow - their optimal information structure sometimes aggregates the private information of the firms to produce a noisy signal for each firm. In my model the information a firm receives can only be conditioned on the type draw of the other firm, not on the combination of types.

Vaissman Guinsburg (2020) develops a model closely related to my example of Cournot competition with uncertainty over linear costs. The main difference is that I assume that firms observe their own types, whereas Vaissman Guinsburg assumes that they receive only the information that their chosen information structure generates for their opponent, so that the only way to conceal information from the opponent is for the player not to learn it themselves (or, equivalently, the only way to learn about one’s own costs is to reveal that information to the opponent). The basic result in both cases is full revelation². Vaissman Guinsburg also shows that the full revelation result does not apply when second-stage equilibrium actions are not interior (i.e. when there is a range of posterior beliefs at which a firm will shut down production). This is also the case in my model, though I do not deal with that scenario in this paper and generally assume interior second-stage actions.

Information sharing in oligopoly is another broad corpus of literature, building on the early work of Novshek and Sonnenschein (1982), Vives (1984), and Fried (1984). Raith (1996) summarises much of the prior literature and draws it together under a general model. As my model is concerned with private values, it is most closely connected to those papers that consider sharing of cost information. Fried (1984) gives a full-revelation result for cost information in Cournot competition with homogenous goods, which is the setting I develop as an example in Section 3.2. Gal-Or (1986) extends the Cournot result to differentiated goods, and shows that the opposite outcome of full concealment of costs results under Bertrand competition. Amir, Jin, and Troege (2010) generalise the earlier results of revelation under Cournot and concealment under Bertrand by dropping distributional assumptions made in earlier papers.

As well as uncertainty over the coefficients on a firm’s own action (which includes private cost uncertainty under an oligopoly interpretation), I consider uncertainty over the cross-effect between a player’s action and their opponent’s. To my knowledge there is little work that considers information sharing when the information pertains to the cross-effect. Chokler, Hon-Snir, Kim and Shitovitz (2006) analyze a Cournot setting with cross-effect uncertainty, and consider which firm benefits when one firm is fully informed and the other uninformed,

²Vaissman Guinsburg suggests that this results from a trade-off between concealing information from the opponent and not having access to it for oneself - however, since full revelation is also the result in the Cournot game in this paper, and in the earlier information sharing literature, it seems that this trade-off does not drive the revelation result.

with the result turning on whether the cross-effect is common or private value. Xu (2010) does consider an information sharing choice with cross-effect uncertainty, rising from an uncertain slope of the demand function in a Cournot setting, which then enters into the cross-effect. This is a constrained case³, and also differs from my model in that the uncertain parameter is common value, whereas I restrict attention to private values.

3.2 An example: Cournot with cost uncertainty

I begin by considering a classic Cournot duopoly with homogenous goods. The result, that firms reveal their cost information as a dominant strategy, is in line with Fried (1984) and the related literature and not striking in itself. Unlike that literature, firms here will have access to ‘persuasive’ information structures, but I will show that they will not choose to use them. However, developing this concrete and relatively straightforward example demonstrates and helps to clarify the analytical approach and structural features of the general model, which I set out in the next section.

Assume that firms simultaneously choose their production quantity q_i , have a constant marginal cost c_i , and face a market price given by the following inverse demand function:

$$p = 1 - q_i - q_j$$

Firm payoffs are then given by:

$$u_i(q_i, q_j, c_i) = (1 - c_i)q_i - q_i q_j - q_i^2$$

Now suppose marginal costs can take on one of two values, $c_l < c_h < 1$, with $c_h < \frac{1+c_l}{2}$ (the latter assumption ensures that firms do choose non-zero quantities). The firms’ costs are drawn i.i.d. and there is a common prior, $P(c_l) = \hat{\mu}$. The firms play a two-stage game with the following timing, acting to maximise their expected payoffs:

1. Each firm simultaneously commits to and announces a **signal structure**, a stochastic mapping from realisations of its own cost into messages sent to the other firm.
2. Costs are drawn. Each firm observes its own draw, a message about the cost of the other firm generated by the other firm’s chosen signal structure, and the message sent from its own signal structure to the other firm. Firms update their beliefs about the cost of the other firm according to Bayes’ rule.
3. Firms simultaneously choose quantities and receive their payoffs.

³Constrained in that it imposes a specific relationship between the value of the cross-effect and the value of the coefficient on the square of a player’s own action.

I will describe the choice of signal structure in more detail shortly, but it will be useful to begin with the final step, the second stage of the game in which firms choose quantities. At this point, firms know their own cost and have some posterior belief about their opponent's cost, $P(c_j = c_l) = \mu_i$. They also know their opponent's posterior about their own cost, μ_j . If firm j 's strategy is given by a function $q_j(\mu_j, \mu_i, c_j)$, then firm i 's problem is:

$$\max_{q_i} E_{\mu_i}(u_i(q_i, q_j(\mu_j, \mu_i, c_j), c_i))$$

Note that:

$$E_{\mu_i}(u_i(q_i, q_j(\mu_j, \mu_i, c_j), c_i)) = (1 - c_i)q_i - q_i E_{\mu_i}(q_j(\mu_j, \mu_i, c_l)) - q_i^2$$

So firm i 's best-response is given by:

$$q_i(\mu_i, \mu_j, c_i) = \frac{1 - c_i - E_{\mu_i}(q_j(\mu_j, \mu_i, c_l))}{2}$$

Substituting and solving gives the equilibrium second-stage quantities as functions of the two firm's posterior expectations about each other's costs:

$$q_i^*(\mu_i, \mu_j, c_i) = \frac{2 - 3c_i + 2E_{\mu_i}(c_j) - E_{\mu_j}(c_i)}{6}$$

Denote by $v_i(\mu_i, \mu_j, c_i)$ the interim expected payoff of firm i when its realised cost is c_i and the posteriors are μ_i and μ_j , given that the firms will play the equilibrium second-stage strategies:

$$v_i(\mu_i, \mu_j, c_i) = E_{\mu_i}(u_i(q_i^*(\mu_i, \mu_j, c_i), q_j^*(\mu_j, \mu_i, c_j), c_i))$$

I now turn to the choice of signal structure. A signal structure is a set of messages and a stochastic mapping that for each cost and message gives the probability that the given message is sent conditional on the given cost being realised. From a firm's point of view, all that matters in the choice of signal structure is the distribution of second-stage payoffs it induces. Second-stage outcomes are determined by the two posterior beliefs and the firm's own cost draw. Note the following:

- The distribution of firm i 's own posterior belief μ_i is entirely determined by the opponent's strategy and is independent of the firm's own choice of signal structure.
- Given that firm j is a rational Bayesian and understands firm i 's chosen signal structure, the distribution of c_i conditional on firm j receiving a given message is one and the same as firm j 's posterior belief after receiving that message. That is, if firm i chooses a structure containing a message that results in firm j assigning probability $\frac{3}{4}$ to $c_i = c_l$, then it must be the case from firm i 's *ex ante* point of view that the probability of having low costs when that message is sent is $\frac{3}{4}$.

Given the above, specifying a posterior belief for firm j fully determines the resulting distribution of second-stage outcomes (holding constant firm j 's own choice of signal structure).

This means that it is without loss to identify messages with the posterior belief μ_j that they induce and, following Kamenica and Gentzkow (2011), firm i 's choice of signal structure can be seen as choosing a distribution τ_i over firm j 's posterior μ_j , subject only to the condition that the distribution be 'Bayes-plausible':

$$E_{\tau_i}(\mu_j) = \hat{\mu}$$

Any given μ_j has an *ex ante* value to firm i , given the opponent's signal structure τ_j : firm i 's expected payoff conditional on firm j having received a message that induced μ_j . Denote this value by $V_i(\mu_j; \tau_j)$:

$$V_i(\mu_j; \tau_j) = E_{\tau_j} \left(E_{\mu_j} (u_i(\mu_i, \mu_j, c_i)) \right)$$

Note that μ_j plays a dual role in this expression. It factors into the function $u_i(\mu_i, \mu_j, c_i)$ as the opponent's posterior, determining firm j 's expected action. However, when it appears in the expectation operator E_{μ_j} , it is acting as the conditional distribution on c_i .

The core insight of Kamenica and Gentzkow (2011) is that once a value is assigned to a receiver's posterior belief in this way, the sender's optimal choice of signal structure is closely related to the *concave closure* \hat{V}_i of V_i , the smallest concave function that is everywhere weakly greater than V_i . In particular, the sender benefits from an information structure that reveals some information (i.e. that sometimes induces posterior beliefs that differ from the prior) if and only if the concave closure is greater at the prior, $\hat{V}_i(\hat{\mu}; \tau_j) > V_i(\hat{\mu}; \tau_j)$. The concave closure also illustrates the support of optimal signal structures (i.e. optimal distributions of posteriors). On any region on which \hat{V}_i diverges from V_i , it forms a convex combination of points of V_i . For the region containing the prior, those points indicate the posteriors that support an optimal signal, as any Bayes-plausible mix of them will yield an expected payoff of $\hat{V}_i(\hat{\mu}; \tau_j)$.

Useful special cases arise when the original function V_i is concave or convex in μ_j . In the former case $\hat{V}_i = V_i$ and an entirely uninformative signal structure is optimal. In the latter case \hat{V}_i coincides with V_i only at the endpoints of its domain - that is, at certainty about the true cost, implying that a fully revealing signal structure is optimal.

Returning to the specifics of the current setting, consider the first and second derivatives of V_i :

$$\begin{aligned} \frac{dV_i(\mu_j; \tau_j)}{d\mu_j} &= E_{\tau_j} \left((v_i(\mu_i, \mu_j, c_l) - v_i(\mu_i, \mu_j, c_h)) + E_{\mu_j} \left(\frac{\partial v_i(\mu_i, \mu_j, c_i)}{\partial \mu_j} \right) \right) \\ \frac{d^2 V_i(\mu_j; \tau_j)}{d\mu_j^2} &= E_{\tau_j} \left(2 \left(\frac{\partial v_i(\mu_i, \mu_j, c_l)}{\partial \mu_j} - \frac{\partial v_i(\mu_i, \mu_j, c_h)}{\partial \mu_j} \right) + E_{\mu_j} \left(\frac{\partial^2 v_i(\mu_i, \mu_j, c_i)}{\partial \mu_j^2} \right) \right) \end{aligned}$$

The dual role of μ_j in firm i 's expected payoff is represented by two distinct effects in the derivatives. The first, difference term in each derivative results from the shift in the conditional probability of firm i having cost c_l that a change in μ_l represents. This is the **own-type effect**. The second term represents the expected changes to firm i 's payoff as equilibrium production levels shift with μ_j . This is the **equilibrium-action effect**.

It is illuminating to consider the second derivative of V_i in terms of the firms' second-stage actions. Note that the envelope theorem applies. That is, $\frac{\partial u_i}{\partial q_i}(q_i^*, q_j^*, c_i) = 0$, so $\frac{d^2 V_i}{d\mu_j^2}(\mu_j; \tau_j)$ can be written in terms of the derivatives of u_i with respect to the other firm's action, q_j . Further, the equilibrium second-stage actions q_i^* and q_j^* are linear in the posterior beliefs, so $\frac{\partial^2 q_j^*}{\partial \mu_j^2} = 0$. It follows that:

$$\begin{aligned} \frac{d^2 V_i(\mu_j; \tau_j)}{d\mu_j^2} = E_{\tau_j} \left[E_{\mu_i} \left(2 \frac{\partial q_j^*}{\partial \mu_j} \left(\frac{\partial u_i}{\partial q_j}(q_i^*, q_j^*, c_i) - \frac{\partial u_i}{\partial q_j}(q_i^*, q_j^*, c_h) \right) \right. \right. \\ \left. \left. + E_{\mu_j} \left(\frac{\partial^2 u_i}{\partial q_i \partial q_j}(q_i^*, q_j^*, c_i) \frac{\partial q_j^*}{\partial \mu_j} \frac{\partial q_i^*}{\partial \mu_j} \right) \right) \right] \end{aligned}$$

Note that $\frac{\partial q_j^*}{\partial \mu_j}$ is negative (an increased probability that firm i has low costs causes firm j to reduce output) and $\frac{\partial u_i}{\partial q_j} = -q_i^*$ is increasing in c_i (when firm i has low costs, its output is higher, so the negative impact of an increase in firm j 's production has greater magnitude). So the own-type effect is positive. As for the equilibrium-action effect, here it consists only of an 'adjustment' effect (if actions were not linear in beliefs it would be more complicated, as later examples will demonstrate). Firm i responds to the increase in firm j 's belief and resulting decrease in firm j 's output by increasing its own output ($\frac{\partial q_i^*}{\partial \mu_j} > 0$), which in turn increases the benefit to firm i of further increases in firm j 's belief:

$$\frac{\partial^2 u_i}{\partial q_i \partial q_j}(q_i^*, q_j^*, c_i) \frac{\partial q_j^*}{\partial \mu_j} = -\frac{\partial q_j^*}{\partial \mu_j} > 0$$

Since both effects are positive, clearly V_i is convex in μ_j regardless of firm j 's choice of signal structure τ_j , and it is a dominant strategy for firm i to choose a completely revealing information structure. The intuition is reasonably straightforward. Firm i always benefits from an increase in firm j 's belief that c_i is low. This benefit is increasing in firm j 's belief, because firm i observes the information firm j receives and can adjust its own output upwards accordingly. The *ex ante* expectation of the benefit is increasing in the conditional probability that firm i actually does draw low costs, because in that case firm i 's output will be higher and it will benefit more from a reduction in firm j 's output. So firm i is willing to allow firm j to learn that its costs are high when it draws high costs, in order to have firm j learn that its costs are low the rest of the time.

3.3 General Model

Environment: There are two players, a set of payoff types $\Theta = \{1, 2\}$, and a common prior $\hat{\mu} \in \Delta\Theta$. Each player independently draws a payoff type θ_i , but the types are initially unknown to either player. Each player i chooses a signal structure τ_i which determines how the other player, j , learns about player i 's payoff type. Each player also chooses an action a_i from a closed interval $A = [0, \bar{A}]$.

Payoffs are given by a function $u : A^2 \times \Theta \rightarrow \mathbb{R}$. When I use this function to represent a given player's payoffs, I will typically use a subscript to indicate which player's payoffs it represents; for example, $u_i(a_i, a_j, \theta_i)$ for the payoffs of player i when they draw payoff type θ_i and take action a_i , and their opponent takes action a_j . This is purely for notational clarity, and should not be taken to indicate that, say, u_i and u_j represent different functions - players are *ex ante* identical and $u_i(a, a', \theta) = u_j(a, a', \theta)$ for any a and a' in A and any θ in Θ .

Signal Structures: Player i 's choice of signal structure τ_i determines the information that player j receives about player i 's payoff type. There is a finite set of messages M , $\Theta \subseteq M$. A signal structure $\tau_i : \Theta \rightarrow \Delta M$ that selects a probability distribution over the set of messages for each of the possible payoff types player i could draw. Denote the set of all possible signal structures by T . Note that T is the Cartesian square of the $|M|$ -simplex.

Timing: The players participate in a game G with the following structure:

1. **First stage:** The players choose signal structures. The choices are simultaneous and the chosen signal structure is observed by the other player.
2. The players' payoff types are independently drawn. Each player observes their own payoff type, and the messages generated by each signal structure.
3. The players update their beliefs about the other player's payoff type using Bayes' rule, based on the message they received and the properties of their opponent's signal structure. The resulting posterior belief of player i is denoted μ_i .
4. **Second stage:** The players choose their actions simultaneously and receive their payoffs.

Strategies: The players make their second-stage action choice having observed the choice of signal structures, their own type draw, and the messages generated by the two signal structures⁴. A **second-stage strategy** for player i is a function $s_{i2} : T^2 \times M^2 \times \Theta \rightarrow A$. I write a second-stage strategy as $s_{i2}(\tau_i, \tau_j, m_i, m_j, \theta_i)$, where m_i is the message player i receives (generated by the signal structure τ_j). Denote the set of all such functions by S_2 . A **first-stage strategy** s_{i1} is an element of ΔT . A strategy s_i for the entire game G is an element of $\Delta T \times S_2$.

Equilibrium: The solution concept is perfect Bayesian equilibrium. A pair of strategies (s_i, s_j) is an equilibrium of G if and only if each s_i maximises player i 's *ex ante* expected utility given s_j (and likewise for player j), and sequential rationality holds for the second-stage strategies s_{i2} and s_{j2} . If $\mu_i(\tau_j, m_i)$ is the posterior belief of player i having received message m_i from signal structure τ_j , then sequential rationality for a second-stage strategy s_{i2} requires that:

$$\forall a_i \in \text{supp } s_{i2}(\tau_i, \tau_j, m_i, m_j, \theta_i), a_i \in \arg \max_{a \in A} E_{\mu_i(\tau_j, m_i)} \left[E_{s_{j2}(\tau_j, \tau_i, m_j, m_i, \theta_j)} (u_i(a, a_j, \theta_i)) \right]$$

⁴The message player j receives about player i 's type is relevant to player i 's action because it informs player i of player j 's posterior belief.

Following the terminology of Koessler et al. (2020), define a strategy s_i as **Markovian** if the second-stage strategies included in s_i do not depend on the messages received beyond the posterior beliefs they induce about payoff types. Specifically, if $\mu_i(\tau_j, m_i) = \mu_i(\tau'_j, m'_i)$ and $\mu_j(\tau_i, m_j) = \mu_j(\tau'_i, m'_j)$ then $s_{i2}(\tau_j, \tau_i, m_i, m_j, \theta_i) = s_{i2}(\tau'_j, \tau'_i, m'_i, m'_j, \theta_i)$. In this case s_{i2} can be written as $s_{i2}(\mu_i, \mu_j, \theta_i)$. A strategy profile or equilibrium is Markovian if the strategies it consists of are Markovian.

For Markovian strategy profiles, the sequential rationality condition can be rewritten in terms of the posterior beliefs:

$$\forall a_i \in \text{supp } s_{i2}(\tau_i, \tau_j, m_i, m_j, \theta_i), a_i \in \arg \max_{a \in A} E_{\mu_i} \left[E_{s_{j2}(\mu_j, \mu_i, \theta_j)}(u_i(a, a_j, \theta_i)) \right] \quad (3.1)$$

Markovian second-stage strategy profiles that meet this condition can be thought of as specifying Bayes-Nash equilibria for each of a set of **second-stage games**, where a second-stage game $G_s(\mu_i, \mu_j)$ is a Bayesian game in which the action space is A , payoffs are given by u , players learn their own payoff type before acting, and the common prior on the payoff types is (μ_i, μ_j) .

Proposition 1. *If u is continuous in a_i and a_j and there exist Markovian second-stage strategies s_{i2} and s_{j2} such that:*

- *the sequential rationality condition (3.1) is satisfied for s_{i2} given s_{j2} and vice versa*
- *s_{i2} and s_{j2} are continuous in the posteriors μ_i and μ_j*

then G has a Markovian equilibrium in which the second-stage strategies are s_{i2} and s_{j2} .

Proof. Consider the reduced first-stage game $\bar{G}(s_{i2}, s_{j2})$ that results from fixing the (Markovian) second-stage strategies s_{i2} and s_{j2} and allowing the players to choose first-stage strategies. Clearly if s_{i2} and s_{j2} satisfy sequential rationality, any equilibrium of \bar{G} is a Markovian equilibrium of G . The posterior induced by a message from player i 's signal structure, $\mu_j(\tau_i, m_j)$, is continuous in τ_j (except where $P(m_j) = 0$). If u is continuous in actions and s_{i2} and s_{j2} are continuous in the posterior beliefs, then it follows that the expected payoffs of the players are continuous in τ_i and τ_j . Since T is compact and convex, $\bar{G}(s_{i2}, s_{j2})$ has a mixed-strategy equilibrium by the existence theorem of Glicksberg (1952). \square

Proposition 1 essentially states that if there exist equilibria for all the possible second-stage games which are continuously related (in posterior beliefs), then there exists a Markovian equilibrium for the overall game. For payoff structures that meet this condition, this demonstrates existence without relying on the possibility of correlated equilibrium as in Koessler et al. (2020). Games in which a Markovian equilibrium exists are also of particular interest, because focusing on such equilibria ensures that the first-stage strategies depend only on the effects of communicating information about the player's payoff type (in a Markovian equilibrium, the first-stage choice of signal structure is not serving to provide a correlating or equilibrium-selection device for second-stage play). In the next section, I consider a class of payoff structures that ensure the conditions of Proposition 1 are met.

3.3.1 Second-stage games with belief-continuous equilibria

I will now focus on a class of second-stage games characterised by three assumptions on the payoff function u :

Assumption 1 (concavity): Payoffs are strictly concave in a player's own action: for any a_i and a_j in A and any θ_i in Θ , $\frac{\partial^2 u_i}{\partial a_i^2}(a_i, a_j, \theta_i) < 0$.

Assumption 2 (partial cross-effect linearity): For any θ_i in Θ , the cross-effect $\frac{\partial^2 u_i}{\partial a_i \partial a_j}$ does not vary in a_j :

$$\frac{\partial^3 u_i(a_i, a_j, \theta_i)}{\partial a_i \partial a_j^2} = 0$$

Assumption 3 (own-action dominance): The own-action second-derivative of payoffs at some action pair is larger in magnitude than the cross-effect. That is, there exists $k \in (0, 1)$ such that for any a_i and a_j in A and any θ_i in Θ :

$$\left| \frac{\frac{\partial^2 u_i}{\partial a_i \partial a_j}(a_i, a_j, \theta_i)}{\frac{\partial^2 u_i}{\partial a_i^2}(a_i, a_j, \theta_i)} \right| < k$$

Assumptions 1 and 3 yield a unique equilibrium for a game of complete information (such as the second-stage games with both posteriors equal to 0 or 1, in which payoff types are common knowledge), by a standard contraction mapping result⁵. The following results demonstrate that along with assumption 2, they are sufficient to guarantee unique and belief-continuous equilibria for all second-stage games, ensuring that a Markovian equilibrium of the over-all game exists and allowing analysis without concerns over the selection of second-stage equilibria.

Lemma 1. *If the payoff function u satisfies Assumptions 1-3, then for any pair of posterior beliefs (μ_i, μ_j) the second-stage game $G_S(\mu_i, \mu_j)$ has a unique, pure-strategy equilibrium.*

Proof. Consider some second-stage game $G_S(\mu_i, \mu_j)$. A strategy \hat{a}_i for player i is an action for each possible type draw, θ_i . By Assumption 1 and the compactness of A , given a strategy a_j for player j , there is a unique best-response \hat{a}_i^* :

$$\hat{a}_i^*(\hat{a}_j, \theta_i) = \max_{a_i \in A} E_{\mu_i} u_i(a_i, \hat{a}_j(\theta_j), \theta_i)$$

Applying the implicit function theorem to the first-order condition for interior values of \hat{a}_i^* gives, for any $\theta \in \Theta$:

$$\frac{\partial \hat{a}_i^*(\hat{a}_j, \theta_i)}{\partial \hat{a}_j(\theta)} = - \frac{\mu_i(\theta) \frac{\partial^2 u_i(\hat{a}_i^*(\hat{a}_j, \theta_i), \hat{a}_j(\theta), \theta_i)}{\partial a_i \partial a_j}}{E_{\mu_i} \left(\frac{\partial^2 u_i(\hat{a}_i^*(\hat{a}_j, \theta_i), \hat{a}_j(\theta_j), \theta_i)}{\partial a_i^2} \right)}$$

⁵See e.g. chapter 2 of Vives (1999)

Take a strategy profile \hat{a} to be a point in the metric space (A^4, d_∞) where d_∞ is the maximum metric:

$$d_\infty(\hat{a}, \hat{a}') = \max_{p \in \{i, j\}, \theta \in \Theta} |\hat{a}_p(\theta) - \hat{a}'_p(\theta)|$$

Consider any two strategies \hat{a} and \hat{a}' . Define:

$$\hat{a}_t = \frac{t\hat{a} + (1-t)\hat{a}'}{d_\infty(\hat{a}, \hat{a}')}, t \in [0, 1]$$

For any $\theta \in \Theta$, this gives:

$$\left| \frac{\partial \hat{a}_t(\theta)}{\partial t} \right| \leq 1$$

Now consider the best-response mapping \hat{a}_i^* , and denote $\hat{a}_i^*(\hat{a}_t)$ by \hat{a}_t^* . Note that for any $\theta_i \in \Theta$ and interior best-responses:

$$\left| \frac{\partial \hat{a}_t^*(\theta_i)}{\partial t} \right| \leq \sum_{\theta_j \in \Theta} \left| \frac{\partial \hat{a}_t(\theta_j)}{\partial t} \frac{\partial \hat{a}_t^*(\theta_i)}{\partial \hat{a}_t(\theta_j)} \right| \leq \sum_{\theta_j \in \Theta} \left| \frac{\partial \hat{a}_t^*(\theta_i)}{\partial \hat{a}_t(\theta_j)} \right| \leq \frac{E_{\mu_i} \left(\frac{\partial^2 u(\hat{a}_t^*(\theta_i), \hat{a}_t(\theta_j), \theta_i)}{\partial a_i \partial a_j} \right)}{-E_{\mu_i} \left(\frac{\partial^2 u(\hat{a}_t^*(\theta_i), \hat{a}_t(\theta_j), \theta_i)}{\partial a_i^2} \right)} < k \in (0, 1)$$

where the last inequality follows from Assumptions 2 and 3. It follows⁶ that $d_\infty(\hat{a}_i^*(\hat{a}), \hat{a}_i^*(\hat{a}')) < kd_\infty(\hat{a}, \hat{a}')$. This demonstrates that \hat{a}_i^* is a contraction and has a unique fixed point, yielding a unique equilibrium of $G_S(\mu_i, \mu_j)$. \square

Given the existence of a unique equilibrium for each G_S , the equilibrium second stage strategy of player i can be written as a function $a_i^*(\mu_i, \mu_j, \theta_i)$. To simplify notation, I will write a profile of posteriors as $\mu = (\mu_i, \mu_j)$ and simply refer to the resulting equilibrium actions by $a_i^*(\mu, \theta_i)$ and $a_j^*(\mu, \theta_j)$.

Lemma 2. *If the payoff function u satisfies Assumptions 1-3, the equilibrium actions $a_i^*(\mu, \theta_i)$ and $a_j^*(\mu, \theta_j)$ are continuous in μ_i and μ_j , and when interior, they are differentiable in μ_i and μ_j . Consequently, a Markovian equilibrium of the whole game exists by Proposition 1.*

Proof. When equilibrium strategies are interior to A , they must satisfy the first-order condition:

$$E_{\mu_i} \left(\frac{\partial u_i}{\partial a_i} (a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j), \theta_i) \right) = 0$$

Consider a function $G : A^4 \times [0, 1]^2 \rightarrow \mathbb{R}^4$ giving the own-action first derivative of expected utility (i.e. the left-hand side of the above first-order condition) for each player-type pair. By the implicit function theorem, the equilibrium actions are given by functions differentiable

⁶This does not account for corner solutions to player i 's problem at which the first-order condition does not apply ($a_i^* = \bar{A}$ or $a_i^* = 0$). Similar arguments apply when this is the case - note that while $\frac{\partial \hat{a}_i^*(\hat{a}_j, \theta_i)}{\partial \hat{a}_j(\theta)}$ does not always exist, left and right derivatives do exist, and are either zero or given by the interior expression stated in the proof.

in μ_i and μ_j if the Jacobian of G with respect to the actions, $D_a G$, is invertible. The determinant of the Jacobian is given by:

$$|D_a G| = 1 - E_{\mu_i, \mu_j} \left(\frac{\frac{\partial^2 u_i(a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j), \theta_i)}{\partial a_i \partial a_j} \frac{\partial^2 u_j(a_j^*(\mu, \theta_j), a_i^*(\mu, \theta_i), \theta_j)}{\partial a_i \partial a_j}}{E_{\mu_i} \left(\frac{\partial^2 u_i(a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j))}{\partial a_i^2} \right) E_{\mu_j} \left(\frac{\partial^2 u_j(a_j^*(\mu, \theta_j), a_i^*(\mu, \theta_i))}{\partial a_j^2} \right)} \right)$$

This is positive by Assumptions 2 and 3.

This demonstrates that equilibrium actions are given by differentiable functions when the actions fall on the interior of A . On a region where some actions fall on the interior of A and some do not, an analogous argument applies to show that the interior actions are still given by differentiable functions⁷. Equilibrium actions will not typically be differentiable at a point where one or more equilibrium actions reach the boundary of A (that is, points at which an equilibrium action falls on the boundary but the first-order condition is still satisfied), but they will be continuous at such points as they are given by differentiable functions in a neighbourhood around the point, which must have the same limit at that point (the unique equilibrium). \square

The consequence of this result is that the first-stage information design problem in games that meet Assumptions 1-3 can be analyzed by considering the derivative of a player's expected payoff conditional on managing to induce a particular belief in their opponent, as this expected payoff is differentiable in beliefs when the equilibrium actions are.

3.3.2 First-stage best responses and the ex-ante value of posteriors

Consider player i 's best response to a Markovian strategy $s_j = (s_{j1}, s_{j2})$. Define player i 's **interim value function** \bar{V}_i and **ex-ante value function** V_i as follows:

$$\bar{V}_i(\mu_j; \mu_i, s_{j2}) = E_{\mu_j} \max_{a_i \in A} \left(E_{\mu_i} E_{s_{j2}(\mu_i, \mu_j, \theta_j)} u_i(a_i, a_j, \theta_i) \right)$$

$$V_i(\mu_j; s_j) = E_{s_{j1}} \left(\bar{V}_i(\mu_j; \mu_i(\tau_j, m_j), s_{j2}) \right)$$

The ex-ante value function gives player i 's expected payoff conditional on sending a message that induces the posterior μ_j , given that the opponent's strategy is s_j . This is an expectation over the interim value function, which additionally conditions player i 's expected payoff on the value of their own posterior, μ_i . Note that these definitions rely on the fact that the distribution of θ_i conditional on sending a μ_j -inducing message is simply μ_j ; hence the expectations with respect to μ_i and μ_j in the above expressions.

⁷The preceding argument can be applied to generate differentiable functions that satisfy only first-order conditions for player-type pairs with interior equilibrium actions, while fixing the other actions at \bar{A} or 0.

Player i 's expected payoff given a first-stage strategy choice is simply the resulting expectation over V_i . That is, first-stage best responses to s_j , $\hat{s}_i(s_j)$, are given by:

$$\hat{s}_i(s_j) = \arg \max_{s_i \in \Delta T} E_{s_i}(V_i(\mu_j(\tau_i, m_j); s_j))$$

Lemma 3. *Denote by $\hat{V}_i(\mu_j; s_j)$ the concave closure of V_i in μ_j , as defined in Kamenica and Gentzkow (2011). If $V_i(\mu_j; s_j)$ is continuous in μ_j , the expected payoff of player i when best-responding to a Markovian strategy s_j is given by the value of \hat{V}_i at the prior, $\hat{\mu}$. That is:*

$$\max_{s_i \in \Delta T} E_{s_i}(V_i(\mu_j(\tau_i, m_j); s_j)) = \hat{V}_i(\hat{\mu}; s_j)$$

Proof. Given that V_i assigns a value to each μ_j player i can induce, and that continuity implies upper semicontinuity, player i 's problem can be taken as an instance of the sender's problem in Kamenica and Gentzkow (2011). Equivalent results follow. \square

Since Lemma 2 implies the continuity of V_i for games that satisfy Assumptions 1-3, Lemma 3 demonstrates that the problem of best-responding in the first stage of such games can be treated as a Bayesian Persuasion problem, which allows for the application of other useful results from Kamenica and Gentzkow (2011).

It will be useful to consider benchmark information structures that either fully reveal a player's type, or disclose no information beyond the prior. Denote by \bar{s}_0 a **fully-concealing information structure** under which $P(m = 1|\theta = 1) = 1$ and $P(m = 1|\theta = 2) = 1$ (the posterior induced by this structure is always the prior, $\hat{\mu}$). Denote by \bar{s}_1 a **fully-revealing information structure** under which $P(m = 1|\theta = 1) = 1$ and $P(m = 2|\theta = 2) = 1$ (this structure reveals the player's type to the opponent, leading to a posterior of either 0 or 1).

Corollary 3. *If $V_i(\mu_j; s_j)$ is convex in μ_j , $\bar{s}_1 \in \hat{s}_i(s_j)$, with $\hat{s}_i(s_j) = \{\bar{s}_1\}$ if it is strictly convex. Similarly, if $V_i(\mu_j; s_j)$ is concave in μ_j , $\bar{s}_0 \in \hat{s}_i(s_j)$, with $\hat{s}_i(s_j) = \{\bar{s}_0\}$ if it is strictly concave.*

First-stage equilibria for a given game can be sought by investigating the ex-ante value function, and the underlying interim value function. In particular, a fully-concealing equilibrium in which both players choose \bar{s}_0 in the first stage will exist if there exist sequentially rational second-stage strategies (s_{i2}, s_{j2}) such that $\bar{V}_i(\mu_j; \hat{\mu}, s_{j2})$ and $\bar{V}_j(\mu_i; \hat{\mu}, s_{i2})$ are both concave. Similarly, a fully-revealing equilibrium will exist if there exist sequentially rational second-stage strategies (s_{i2}, s_{j2}) such that $\hat{\mu}\bar{V}_i(\mu_j; 1, s_{j2}) + (1 - \hat{\mu})\bar{V}_i(\mu_j; 0, s_{j2})$ and $\hat{\mu}\bar{V}_j(\mu_i; 1, s_{i2}) + (1 - \hat{\mu})\bar{V}_j(\mu_i; 0, s_{i2})$ are both convex. This investigation is most straightforward when \bar{V}_i is differentiable in μ_j (which implies that V_i is as well), and convexity or concavity can be established based on the sign of $\frac{\partial^2 V_i}{\partial \mu_j^2}$. Differentiability follows when the second-stage strategies are differentiable; as discussed in Section 3.3.1, for games that meet Assumptions 1-3 this is the case for the unique second-stage equilibrium actions, except at points at which they reach the boundaries of A .

In such cases, the discussion of the second-derivative in the Cournot case in Section 3.2 generalises, and the direct and own-type effects identified for that application can be identified in the more general case. When considering games that meet Assumptions 1-3, the unique second-stage equilibria can be taken as fixed when considering the first stage, and the interim value function can be written as:

$$\bar{V}_i(\mu_j; \mu_i) = E_{\mu_i, \mu_j} \left(u_i(a_i^*(\mu_i, \mu_j, \theta_i), a_j^*(\mu_j, \mu_i, \theta_j), \theta_i) \right)$$

The second derivative is:

$$\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} = D_i(\mu_i, \mu_j) + I_i(\mu_i, \mu_j)$$

where $D(\mu_i, \mu_j)$ is the **equilibrium-action effect** and $I(\mu_i, \mu_j)$ is the **own-type effect**. The equilibrium-action effect is given by:

$$\begin{aligned} D_i(\mu_i, \mu_j) = E_{\mu_i} E_{\mu_j} & \left(\frac{\partial u_i(a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j), \theta_i)}{\partial a_j} \frac{\partial^2 a_j^*(\mu, \theta_j)}{\partial \mu_j^2} \right. \\ & + \frac{\partial^2 u_i(a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j), \theta_i)}{\partial a_j^2} \left(\frac{\partial a_j^*(\mu, \theta_j)}{\partial \mu_j} \right)^2 \\ & \left. + \frac{\partial^2 u_i(a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j), \theta_i)}{\partial a_i \partial a_j} \frac{\partial a_j^*(\mu, \theta_j)}{\partial \mu_j} \frac{\partial a_i^*(\mu, \theta_i)}{\partial \mu_j} \right) \end{aligned}$$

This effect is the second-order impact on player i 's expected payoff of a shift in player j 's beliefs about player i 's payoff type, holding constant the conditional distribution of player i 's type. If player i could manipulate player j 's belief through some deceptive means (if they could increase player j 's belief that they drew type 1 given a particular message without actually altering the information structure determining when that message would be sent), then this would be the change in the benefit they derive as μ_j increases. The first two terms give the second-order effect of the change in player j 's action on player i 's expected payoff, and the third term is the impact that the shift in player i 's equilibrium action has on the first-order effect of a change in player j 's action.

The own-type effect is given by:

$$I_i(\mu_i, \mu_j) = 2E_{\mu_i} \left(\frac{\partial a_j^*(\mu, \theta_j)}{\partial \mu_j} \left(\frac{\partial u_i(a_i^*(\mu, 1), a_j^*(\mu, \theta_j), 1)}{\partial a_j} - \frac{\partial u_i(a_i^*(\mu, 2), a_j^*(\mu, \theta_j), 2)}{\partial a_j} \right) \right)$$

This effect is the second-order impact on player i 's expected payoff of a shift in the conditional distribution of their type. As μ_j increases, a_j^* changes, which has an impact on player i 's expected payoff - however, the magnitude and direction of this impact, $\frac{\partial u_i}{\partial a_j}$ will (typically) vary depending on player i 's type, both directly through a different type representing a different payoff structure, and indirectly in that $\frac{\partial u_i}{\partial a_j}$ may depend on player i 's own choice of action a_i , which will depend on their payoff type. The change in the expectation of $\frac{\partial u_i}{\partial a_j}$ as μ_j shifts is a second-order effect on player i 's ex-ante expected payoff.

The definitions of these effects show that optimal choices of information structure in the first stage depend on the structure of the payoff function u for each of the possible type draws (that is, on the various derivatives of u), and on the way that equilibrium actions respond to changes in the posterior beliefs (the derivatives of a_i^* and a_j^* with respect to μ_j). In principle, the latter can be expressed in terms of the former for any game meeting the assumptions of section 3.3.1. In the next section, I will focus on a class of games in which doing so is relatively straightforward, and demonstrate how different forms of uncertainty lead to various kinds of first-stage equilibria.

3.4 Quadratic games

Consider a payoff function of the following form:

$$u_i(a_i, a_j, \theta) = z_{\theta_i} a_i - \frac{1}{2} y_{\theta_i} a_i^2 + x_{\theta_i} a_i a_j + H(a_j, \theta_i) \quad z_{\theta_i} > 0, y_{\theta_i} > 0$$

Here z_{θ_i} , y_{θ_i} , and x_{θ_i} are constants for a given payoff type θ_i , and do not vary with actions, while $H(a_j, \theta_i)$ is a function that does not depend on a_i and is twice-differentiable in a_j . The H function can be thought of as an externality that player j exerts on player i , an impact of a_j on player i 's payoff that is independent of player i 's choice of action. I will refer to a game in which payoffs take this form as a **quadratic game**.

In a quadratic game, Assumption 3 is considerably simplified, and reduces to:

$$\left| \frac{x_{\theta_i}}{y_{\theta_i}} \right| < 1$$

Quadratic games satisfying this assumption are particularly tractable in that equilibrium actions (when interior) and their derivatives can be solved for and an explicit expression for $\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2}$ in terms of the parameters can be derived, as in the Cournot example of Section 3.2, which is a quadratic game. In doing so, it will be useful to define notation for ratios of the relevant derivatives:

$$\rho_{\theta_i} = \frac{x_{\theta_i}}{y_{\theta_i}}, \quad \gamma_{\theta_i} = \frac{z_{\theta_i}}{y_{\theta_i}}$$

Additionally, I will denote by θ'_i the payoff type that player i did *not* draw. That is, if $\theta_i = 1$, $\theta'_i = 2$, and vice versa.

The first-order conditions that apply to interior equilibrium actions in a quadratic game are of the form:

$$a_i^*(\mu, \theta_i) = \gamma_{\theta_i} + \rho_{\theta_i} E_{\mu_i}(a_j^*(\mu, \theta_j))$$

Solving the resulting system of equations gives:

$$a_i^*(\mu, \theta_i) = \frac{\gamma_{\theta_i} + \rho_{\theta_i} E_{\mu_i}(\gamma_{\theta_j}) + (1 - \mu_j(\theta_i)) E_{\mu_i}(\rho_{\theta_j}) (\rho_{\theta_i} \gamma_{\theta'_i} - \rho_{\theta'_i} \gamma_{\theta_i})}{1 - E_{\mu_i}(\rho_{\theta_j}) E_{\mu_j}(\rho_{\theta_i})}$$

Taking derivatives with respect to μ_j and making some simplifying substitutions gives:

$$\begin{aligned}\frac{\partial a_i^*(\mu, \theta_i)}{\partial \mu_i} &= \frac{\rho_{\theta_i}(a_j^*(\mu, 1) - a_j^*(\mu, 2))}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} \\ \frac{\partial^2 a_i^*(\mu, \theta_i)}{\partial \mu_i^2} &= \frac{2E_{\mu_j}(\rho_{\theta_i})(\rho_1 - \rho_2)}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} \frac{\partial a_i^*(\mu, \theta_i)}{\partial \mu_i} \\ \frac{\partial a_i^*(\mu, \theta_i)}{\partial \mu_j} &= \rho_{\theta_i} E_{\mu_i} \left(\frac{\partial a_j^*(\mu, \theta_j)}{\partial \mu_j} \right)\end{aligned}$$

Noting that for quadratic games $\frac{\partial u_i(a_i, a_j, \theta_i)}{\partial a_j} = x_{\theta_i} a_i + \frac{\partial g(a_j, \theta_i)}{\partial a_j}$, I can now set out $\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2}$ for these games, beginning with the own-type effect:

$$I_i(\mu_i, \mu_j) = 2 \frac{E_{\mu_i}(\rho_{\theta_j})(a_i^*(\mu, 1) - a_i^*(\mu, 2))}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} \left(x_1 a_i^*(\mu, 1) - x_2 a_i^*(\mu, 2) + \frac{\partial H(a_j, 1)}{\partial a_j} - \frac{\partial H(a_j, 2)}{\partial a_j} \right)$$

The equilibrium-action effect is:

$$\begin{aligned}D_i(\mu_i, \mu_j) &= \frac{E_{\mu_i}(\rho_{\theta_j})(a_i^*(\mu, 1) - a_i^*(\mu, 2))}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} \left(\frac{2E_{\mu_i}(\rho_{\theta_j})(\rho_1 - \rho_2)}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} E_{\mu_j}(x_{\theta_i} a_i^*(\mu, \theta_i) + \frac{\partial H(a_j, \theta_i)}{\partial a_j}) \right. \\ &\quad \left. + E_{\mu_j}(x_{\theta_i} \rho_{\theta_i}) \frac{E_{\mu_i}(\rho_{\theta_j})(a_i^*(\mu, 1) - a_i^*(\mu, 2))}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} \right)\end{aligned}$$

Making use of these expressions allows for consideration of particular forms of uncertainty, by constraining which parameters vary across the payoff types. I begin with a result that encompasses the full-revelation result described for homogeneous Cournot duopoly with cost uncertainty in Section 3.2.

Proposition 2. *For a quadratic game with interior equilibrium actions for every second-stage game, if the cross-effect does not vary with payoff type ($x_1 = x_2$) and:*

- **either** $y_1 = y_2$ and $H(a_j, 1) \equiv H(a_j, 2)$
- **or** $(z_1 - z_2)(y_1 - y_2) \leq 0$ and $\frac{\partial H(a_j, 1)}{\partial a_j} \equiv \frac{\partial H(a_j, 2)}{\partial a_j} \equiv 0$

then as long as either y or z does vary with payoff type, a fully-revealing signal strictly dominates all first-stage strategies that are not fully-revealing, and consequently payoff types are revealed in any equilibrium.

Proof. Recall that:

$$a_i^*(\mu, \theta_i) = \rho_{\theta_i} E_{\mu_i}(a_j^*(\mu, \theta_j)) + \gamma_{\theta_i}$$

This expression is increasing in z_{θ_i} and decreasing in y_{θ_i} (since $a_i^*(\mu, \theta_i) > 0$ by assumption). It follows that $(z_1 - z_2)(y_1 - y_2) \leq 0$ and one or both of $z_1 \neq z_2$ or $y_1 \neq y_2$ holding implies that $a_i^*(\mu, 1) - a_i^*(\mu, 2)$ is nonzero for any beliefs.

Assume that $y_1 = y_2 = y$ and $H(a_j, 1) \equiv H(a_j, 2)$. In this case $x_1 = x_2 = x$ implies that $\rho_1 = \rho_2$ and that equilibrium actions are linear in beliefs, and the equilibrium-action effect reduces to:

$$x^4 y^{-3} \frac{(a_i^*(\mu, 1) - a_i^*(\mu, 2))^2}{(1 - x^2 y^{-2})^2} > 0$$

The own-type effect similarly reduces to:

$$2x^2 y^{-1} \frac{(a_i^*(\mu, 1) - a_i^*(\mu, 2))^2}{1 - x^2 y^{-2}} > 0$$

Now assume instead that $(z_1 - z_2)(y_1 - y_2) \leq 0$ and $\frac{\partial H(a_j, 1)}{\partial a_j} \equiv \frac{\partial H(a_j, 2)}{\partial a_j} \equiv 0$. Since $(y_1 - y_2)$ may be nonzero, actions may not be linear in beliefs, and the equilibrium-action effect is:

$$\begin{aligned} & \frac{x^4 (E_{\mu_i}(y_{\theta_j}^{-1}))^2}{(1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1}))^2} \left((a_i^*(\mu, 1) - a_i^*(\mu, 2)) (y_1^{-1} - y_2^{-1}) E_{\mu_j}(a_i^*(\mu, \theta_i)) \right. \\ & \quad \left. + (a_i^*(\mu, 1) - a_i^*(\mu, 2))^2 E_{\mu_j}(y_{\theta_i}^{-1}) \right) \end{aligned}$$

Since $(z_1 - z_2)(y_1 - y_2) \leq 0$ implies that $(a_i^*(\mu, 1) - a_i^*(\mu, 2))$ and $(y_1^{-1} - y_2^{-1})$ have the same sign, all terms in the above expression are positive.

The own-type effect is:

$$\frac{2x^2 E_{\mu_i}(y_{\theta_j}^{-1}) (a_i^*(\mu, 1) - a_i^*(\mu, 2))^2}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1})} > 0$$

It follows that under either set of assumptions, $\bar{V}_i(\mu_j; \mu_i)$ is strictly convex in μ_j for any μ_i , and that \bar{s}_1 strictly dominates any first-stage strategy that is not fully revealing (i.e. one that puts positive probability on a message that induces a posterior μ_j not equal to 0 or 1). \square

The Cournot example of Section 3.2 clearly falls under this result, with $H = 0$, $y_1 = y_2 = 2$, $x_1 = x_2 = -1$, $z_1 = 1 - c_l$ and $z_2 = 1 - c_h$. It also follow that the conclusion extends to non-homogenous goods, whether substitutes or complements (other values of x) and to quadratic costs *if* the linear and quadratic components of costs are positively correlated. That is, if the cost function has the form:

$$c_i(q_i) = \alpha_{\theta_i} q_i + \beta_{\theta_i} q_i^2$$

then full revelation will result if whenever the firm has a higher α it also has a higher β , since in this case $z_{\theta_i} = 1 - \alpha_{\theta_i}$ and $y_{\theta_i} = 2(1 + \beta_{\theta_i})$.

However, in a setting in which the linear and quadratic components of costs are negatively correlated (i.e. the two possibilities are that a firm has a marginal cost function with a high

intercept but low gradient, or one with a low intercept but high gradient) Proposition 2 does not apply and full revelation may not be the result. The next result demonstrates that \bar{V} can potentially be concave in such cases.

Proposition 3. *There exists a quadratic game with $(y_1 - y_2)(z_1 - z_2) > 0$, $x_1 = x_2$, and $\frac{\partial H(a_j, 1)}{\partial a_j} \equiv \frac{\partial H(a_j, 2)}{\partial a_j} \equiv 0$ such that for some $\bar{\mu} \in [0, 1]$, $\bar{V}_i(\mu_j; \bar{\mu})$ is strictly concave in μ_j . Consequently, for some prior $\hat{\mu} = \bar{\mu}$ there exists an equilibrium of that game in which both players choose a fully-concealing information structure in the first stage.*

Proof. I will show that a game with the required properties can be constructed by fixing parameters other than z_1 and considering the properties of the resulting games as z_1 varies.

Given $x_1 = x_2 = x$ and $\frac{\partial H(a_j, 1)}{\partial a_j} \equiv \frac{\partial H(a_j, 2)}{\partial a_j} \equiv 0$, combining the direct and own-type effects gives:

$$\begin{aligned} \frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} &= \frac{x^2 E_{\mu_i}(y_{\theta_j}^{-1})(a_i^*(\mu, 1) - a_i^*(\mu, 2))}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1})} \left((a_i^*(\mu, 1) - a_i^*(\mu, 2)) \left(1 + \frac{1}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1})} \right) \right. \\ &\quad \left. + x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(a_i^*(\mu, \theta_i))(y_1^{-1} - y_2^{-1}) \right) \end{aligned}$$

Note that, assuming interior solutions for the moment:

$$\begin{aligned} a_i^*(\mu, 1) - a_i^*(\mu, 2) &= \frac{\gamma_1 - \gamma_2 + x E_{\mu_i}(\gamma_j)(y_1^{-1} - y_2^{-1}) + x^2 E_{\mu_i}(y_{\theta_j}^{-1}) \left(\frac{z_2 - z_1}{y_1 y_2} \right)}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1})} \\ \frac{\partial}{\partial z_1} (a_i^*(\mu, 1) - a_i^*(\mu, 2)) &= \frac{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) y_2^{-1} + \mu_i x (y_1^{-1} - y_2^{-1})}{y_1 (1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1}))} \end{aligned}$$

Assume without loss of generality that $y_1 > y_2$. By Assumption 3, $x^2 E_{\mu_i}(y_{\theta_j}^{-1}) y_2^{-1} < 1$, and it follows that for a given μ_i , there is a value of x below which the above derivative is positive (any negative x will suffice). Alternatively, for any given x there is a value of μ_i below which the derivative is positive.

Note also that if $z_1 = z_2$, then:

$$a_i^*(\mu, 1) - a_i^*(\mu, 2) = \frac{z_2(y_1^{-1} - y_2^{-1})(1 + x E_{\mu_i}(y_{\theta_j}^{-1}))}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1})} < 0$$

Note further that assuming that $x > -\frac{y_2}{y_1}$ (this is sufficient to guarantee a_i^* positive in this case) and that \bar{A} is larger than the maximum a_i^* ensures that these equilibrium actions are interior.

So parameters can be chosen such that the difference in actions $a_i^*(\mu, 1) - a_i^*(\mu, 2)$ is negative when $z_1 = z_2$ and is strictly linearly increasing in z_1 . This implies the existence in such a case, for any given μ_j , of a value $\bar{z} > z_2$ such that when $z_1 = \bar{z}$, $a_i^*(\mu, 1) - a_i^*(\mu, 2) = 0$, as

long as \bar{A} is sufficiently large that the resulting values of a_i^* are less than \bar{A} for any μ (since $a_i^*(\mu, 1)$ is increasing in z_1 and can be assumed positive at $z_1 = z_2$, the lower bound at 0 is not a concern).

In fact, \bar{z} must be the same for all values of μ_j . To see this, note that if player i plays a strategy in which they select the same action regardless of payoff type, player j 's best response does not depend on μ_j . If player j 's action does not vary with μ_j , player i 's best response also does not depend on μ_j . It follows that if such a strategy for player i forms part of an equilibrium for some second-stage game $G_S(\mu_i, \mu_j)$, then that equilibrium is also an equilibrium of any other $G_S(\mu_i, \mu'_j)$ for any $\mu'_j \in [0, 1]$. It then follows from the uniqueness of second-stage equilibria that for a given μ_i there is a unique \bar{z} at which $a_i^*(\mu, 1) - a_i^*(\mu, 2) = 0$ for any μ_j .

Clearly when $z_1 = \bar{z}$, $\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} = 0$. It also gives, for any μ_j :

$$\frac{\partial}{\partial z_1} \frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} = \frac{x^4 E_{\mu_i}(y_{\theta_j}^{-1})^2 E_{\mu_j}(a_i^*(\mu, \theta_i))(y_1^{-1} - y_2^{-1})}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1})} \left(\frac{\partial}{\partial z_1} (a_i^*(\mu, 1) - a_i^*(\mu, 2)) \right) < 0$$

It follows that there exists an ϵ such that when $z_1 \in (\bar{z}, \bar{z} + \epsilon)$, $\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} < 0$ for any μ_j . \square

The proof of Proposition 3 also demonstrates that for games with $(y_1 - y_2)(z_1 - z_2) > 0$, \bar{V}_i is sometime convex and sometimes concave in μ_j ⁸. Why do such games differ from the always-convex cases laid out in Proposition 2?

The difference lies in the equilibrium-action effect. The own-type effect is still always positive in the games described by Proposition 3, because with constant x and no externalities, both player j 's response to a shift in μ_j , $\frac{\partial a_j^*}{\partial \mu_j}$, and the shift in the effect of that response on player i 's payoff, the slope of which is $\frac{\partial u_i(a_i, a_j, 1)}{\partial a_j} - \frac{\partial u_i(a_i, a_j, 2)}{\partial a_j}$, are given their sign by the product of the cross-effect x and the difference in player i 's action across types, $a_i^*(\mu, 1) - a_i^*(\mu, 2)$. If actions are strategic complements, for example, and player i 's action is higher when they are type 1, it follows that both player j 's expected action and the expectation of the benefit to player i of an increase in that action are increasing in μ_j .

The equilibrium-action effect, however, can be negative in games with $(y_1 - y_2)(z_1 - z_2) > 0$, because of the following term:

$$\begin{aligned} E_{\mu_i} E_{\mu_j} \left(\frac{\partial u_i(a_i^*(\mu, \theta_i), a_j^*(\mu, \theta_j), \theta_i)}{\partial a_j} \frac{\partial^2 a_j^*(\mu, \theta_j)}{\partial \mu_j^2} \right) \\ = \left(x E_{\mu_j}(a_i^*(\mu, \theta_i)) \right) \left(\frac{2x^2 E_{\mu_i}(y_{\theta_j}^{-1})(y_1^{-1} - y_2^{-1})}{1 - x^2 E_{\mu_i}(y_{\theta_j}^{-1}) E_{\mu_j}(y_{\theta_i}^{-1})} E_{\mu_i} \left(\frac{\partial a_j^*(\mu, \theta_j)}{\partial \mu_j} \right) \right) \end{aligned}$$

⁸Convex examples can be constructed by noting that for any given y_1, y_2 with $y_1 \neq y_2$ and $x_1 = x_2 = x$ such that the equilibrium actions are interior, \bar{V}_i is strictly convex at $z_1 = z_2$, and therefore also strictly convex when z_1 falls in some neighborhood $[z_2 - \epsilon, z_2 + \epsilon]$.

If, for example, we assume that $y_1 - y_2 > 0$, but $z_1 - z_2$ is sufficiently large that $a_i^*(\mu, 1) - a_i^*(\mu, 2) > 0$, then the two components of the above product will have opposite sign. If it is a game of strategic complements ($x > 0$), the expected effect of an increase in a_j on u_i will be positive but a_j^* will be concave in μ_j , from $\frac{\partial a_j^*(\mu, \theta_j)}{\partial \mu_j} > 0$ and $(y_1^{-1} - y_2^{-1}) < 0$. Similarly, if it is a game of strategic substitutes the expected impact will be negative but a_j will be convex.

The explanation for this lies in the denominator of the expressions for a_j^* and its derivatives, $1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})$. This term can be thought of as a ‘feedback’ effect on the equilibrium actions. When player j ’s belief μ_j shifts, the response in their equilibrium action is not only due to the shift in their belief about which of the two actions $a_i^*(\mu, 1)$ and $a_i^*(\mu, 2)$ player i will take, but also to their understanding that player i observes the shift in μ_j and understands that player j ’s action will change, shifting their own actions accordingly. This amplifies the change in equilibrium actions given a change in belief compared to a situation in which the change was hidden from the other player⁹. Note that if the change in μ_j was hidden from player i , player j ’s best response would change linearly in μ_j .

The feedback effect depends on ρ because ρ represents the responsiveness of a player to a change in the expected action of the other player:

$$\frac{\partial a_i^*}{\partial E(a_j^*)} = \rho_{\theta_i}$$

So in the case, for example, in which $x > 0$, $y_1 > y_2$, and $z_1 > z_2$, a_j^* can be increasing but concave in μ_j because while an increase in μ_j can represent an increase in $E_{\mu_j}(a_i^*)$ when $z_1 - z_2$ is large enough, this effect is itself decreasing in μ_j because as μ_j increases, $E_{\mu_j}(\rho_{\theta_i})$ falls - player j ’s expectation of player i ’s responsiveness to the change in a_j^* is falling, which in turn reduces the rate of increase in a_j^* itself. Proposition 3 demonstrates that there are cases in which this effect outweighs the own-type effect (and the other, positive component of the equilibrium-action effect), leading to a concave \bar{V}_i and an incentive for the players to conceal information.

3.4.1 Externalities

In the cases I have considered up to this point, I have essentially assumed away the ‘externality’ component of the payoff function, $H(a_j, \theta_i)$. In this section I will consider the effects of some straightforward (linear and quadratic) externalities that vary across payoff types, which will also illustrate how linear cost uncertainty in Bertrand duopoly leads firms to conceal information.

⁹It can also subdue the change if $E_{\mu_i}(\rho_{\theta_j})$ and $E_{\mu_j}(\rho_{\theta_i})$ have opposite sign, ruled out here but possible in some cases discussed later in this paper.

Linear Externalities and Bertrand Competition

I begin with linear externalities, of the form:

$$H(a_j, \theta_i) = h_{\theta_i} a_j$$

Externalities do not affect equilibrium actions¹⁰, but they do impact $\frac{\partial^2 \bar{V}}{\partial \mu_j^2}$. Compared to some base game without externalities, a linear externality adds the following term to the own-type effect:

$$2 \frac{E_{\mu_i}(\rho_{\theta_j})(a_i^*(\mu, 1) - a_i^*(\mu, 2))}{1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i})} (h_1 - h_2)$$

and the following term to the equilibrium-action effect:

$$\frac{2E_{\mu_i}(\rho_{\theta_j})^2(a_i^*(\mu, 1) - a_i^*(\mu, 2))(\rho_1 - \rho_2)}{(1 - E_{\mu_i}(\rho_{\theta_j})E_{\mu_j}(\rho_{\theta_i}))^2} E_{\mu_j}(h_{\theta_i})$$

It is clear that for a game with no externalities in which \bar{V} is always concave or convex, leading to equilibria with full revelation or no revelation, there are corresponding games with externalities that reverse those outcomes. The next result formalizes this for the case in which payoff types differ only in z and h ; recall that without externalities (uncertainty over z only) these settings lead to full revelation by Proposition 2.

Proposition 4. *For a quadratic game with linear externalities, interior equilibrium actions for every second-stage game, and $x_1 = x_2$, $y_1 = y_2$, a fully-concealing signal strictly dominates all first-stage strategies that are not fully-concealing if and only if:*

$$\frac{h_2 - h_1}{\rho(z_1 - z_2)} > \frac{1 - \frac{\rho^2}{2}}{1 - \rho^2}$$

Conversely, a fully-revealing signal strictly dominates all first-stage strategies that are not fully-revealing if and only if:

$$\frac{h_2 - h_1}{\rho(z_1 - z_2)} < \frac{1 - \frac{\rho^2}{2}}{1 - \rho^2}$$

Proof. For a game meeting the conditions given in the statement of the proposition, combin-

¹⁰Note that the model does not allow players to choose not to participate and receive an outside option. Externalities would potentially impact such a decision, and for the conclusions of this section to apply the externalities would have to be assumed sufficiently small not to do so. In the Bertrand interpretation I give later in this section, this is not a concern, as in that case a non-negative payoff is always achievable for a firm, so a zero-payoff shutdown option would not have an impact.

ing the direct and own-type effects gives:

$$\begin{aligned}
\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} &= \frac{2\rho(\gamma_1 - \gamma_2)}{1 - \rho^2} \left(x(\gamma_1 - \gamma_2) + h_1 - h_2 \right) + \frac{x\rho^3(\gamma_1 - \gamma_2)^2}{(1 - \rho^2)^2} \\
&= \frac{2\rho^2(z_1 - z_2)^2}{y(1 - \rho^2)} \left(\frac{h_1 - h_2}{\rho(z_1 - z_2)} + 1 + \frac{\rho^2}{2(1 - \rho^2)} \right) \\
&= \frac{2\rho^2(z_1 - z_2)^2}{y(1 - \rho^2)} \left(\frac{1 - \frac{\rho^2}{2}}{1 - \rho^2} - \frac{h_2 - h_1}{\rho(z_1 - z_2)} \right)
\end{aligned}$$

The conditions in the proposition follow straightforwardly from the sign of this expression. \square

Bertrand duopoly with linear demand and uncertainty over marginal costs falls under the scope of this result. Consider two firms in a duopolistic market choosing prices p_i and facing a demand function:

$$q_i = 1 - \alpha p_i - \beta p_j$$

If costs are linear, with marginal cost c_i , the resulting payoff function for firm i is:

$$u_i(p_i, p_j, c_i) = (1 - \alpha p_i - \beta p_j)(p_i - c_i), \quad \alpha > 0, |\beta| \leq \alpha$$

Assuming that marginal costs are uncertain and taking on a value of c_1 or c_2 then yields a quadratic game with:

$$y = 2\alpha \quad x = -\beta \quad z_{\theta_i} = 1 + \alpha c_{\theta_i} \quad h(a_j, \theta_i) = \beta c_{\theta_i}$$

Substituting into the condition for fully-concealing equilibria from Proposition 4 and simplifying gives:

$$\left(\frac{\beta}{2\alpha} \right)^2 < \frac{2}{3}$$

This holds by the assumption that $|\beta| \leq \alpha$, so the unique equilibrium is that both firms conceal their costs.

This matches results in the information sharing literature¹¹, but it also serves to illuminate the properties of this particular case that drive the outcome. Consider a Bertrand duopoly with complementary goods, $\beta > 0$. In terms of second-stage equilibrium outcomes, this is structurally equivalent to a Cournot duopoly with goods that are substitutes; in both cases, actions are strategic substitutes. However, in the Cournot case the marginal effect of firm j 's action on firm i 's payoff depends on firm i 's type only through firm i 's equilibrium action, where in the Bertrand case it depends on marginal revenue, $(p_i - c_i)$, which varies directly with type even if the action p_i is held fixed. An increase in the conditional probability that c_i is high implies a higher expected p_i , leading firm j to reduce p_j , which is to firm i 's benefit - but this benefit is falling as that probability rises, because when firm i draws high costs

¹¹Gal-Or (1985), Raith (1996), and Amir, Jin and Troege (2010) all give results in which concealment of cost information is chosen by firms in Bertrand oligopoly.

their price is higher but their marginal revenue, and consequently the marginal benefit they derive from a fall in p_j , is lower. This leads to a concave \bar{V} and a preference for concealing information.

It is the relationship between a firm's cost and its marginal revenue in equilibrium that is crucial here; note the reliance on the parameter restriction that $|\beta| < \alpha$. This restriction is a reasonable one for the setting (it is typically derived from the assumption that a representative consumer has strictly concave utility), but if there were some structurally similar setting in which it did not apply, the full-concealment result would be reversed if $|\beta| \in (\frac{2}{\sqrt{3}}\alpha, 2\alpha)$.

Quadratic Externalities and Partially-Revealing Equilibria

In the various types of quadratic game I have considered up to this point, equilibria have involved either full concealment or full revelation. I now give an example in which u exhibits a quadratic externality that demonstrates the existence of games in which the unique equilibrium involves neither fully-concealing nor fully-revealing first-stage strategies. Consider a quadratic game with the following payoff function:

$$u_i(a_i, a_j, \theta_i) = z_{\theta_i} a_i - a_i a_j - a_i^2 + k_{\theta_i} a_j^2$$

Assume that $z_2 < z_1 < 2z_2$, which guarantees strictly positive equilibrium actions. Note that in this game $E_{\mu_i}(a_j^*(\mu, \theta_j)) = \frac{1}{3}(2E_{\mu_i}(z_j) - E_{\mu_j}(z_i))$. This gives:

$$\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} = \frac{1}{18}(z_1 - z_2)^2 \left(7 - 8 \frac{k_1 - k_2}{z_1 - z_2} (2E_{\mu_i}(z_j) - E_{\mu_j}(z_i)) + 4E_{\mu_j}(k_i) \right)$$

For $k_1 > k_2$, this implies that \bar{V}_i is convex when:

$$\mu_j > \frac{2}{3} \left(\frac{2E_{\mu_i}(z_j) - E_{\mu_j}(z_i)}{z_1 - z_2} - \frac{7 + 4k_1}{8(k_1 - k_2)} + \frac{1}{2} \right)$$

For $k_1 < k_2$ the inequality is reversed. For any given z_1 and z_2 , it is straightforward to find values of k_1 and k_2 such that this threshold falls in $(0, 1)$, implying that \bar{V}_i is convex for some values of μ_j and concave for others.

However, this is not in itself sufficient to demonstrate that a partially-revealing equilibrium exists. Fully-revealing and fully-concealing signals may still be best responses when the value function a player faces has both convex and concave segments. By Lemma 3, the value of player i 's best response to a Markovian strategy s_j is given by the value at the prior of \hat{V}_i , the concave closure of V_i , where V_i is the expectation of \bar{V}_i given s_j . Assume that there is some value $\bar{\mu}_j$ such that V_i is strictly concave in μ_j for $\mu_j < \bar{\mu}_j$ and strictly convex in μ_j for $\mu_j > \bar{\mu}_j$, as is possible in the $k_1 > k_2$ case. In this case we can write \hat{V}_i piecewise as:

$$\hat{V}_i(\mu_j; s_j) = \begin{cases} V_i(\mu_j; s_j) & \mu_j \leq \mu_j^* \\ \frac{1}{1-\mu_j^*} (V_i(1; s_j) + (1 - \mu_j)V_i(\mu_j^*; s_j)) & \mu_j \geq \mu_j^* \end{cases}$$

where μ_j^* is the solution to the following equation if a solution exists in $[0, 1]$, and is 0 otherwise:

$$\frac{\partial V_i(\mu_j^*; s_j)}{\partial \mu_j} = \frac{V_i(1, s_j) - V_i(\mu_j^*; s_j)}{1 - \mu_j}$$

Recall that the value of a best response to s_j is given by the value of $\hat{V}_i(\hat{\mu}; s_j)$ where $\hat{\mu}$ is the common prior. If $\hat{\mu} < \mu^*$, then $\hat{V}_i(\hat{\mu}; s_j) = V_i(\hat{\mu}; s_j)$ and the best response is a fully-concealing signal. Otherwise, $\hat{V}_i(\hat{\mu}; s_j)$ is achieved through a signal that mixes between posteriors of 1 and μ^* . If $\mu^* = 0$ this is a fully-revealing signal, but if $\mu^* > 0$ it is partially-revealing.

In the case under consideration here, it is possible to solve for μ^* . Assuming $\mu^* > 0$, substituting using the definitions of V_i and \bar{V}_i and rearranging gives:

$$\begin{aligned} E_{sj1} \left(E_{\mu_i} \left(u_i(a_i^*(\mu_i, 1, 1), a_j^*(1, \mu_i, \theta_j), 1) - u_i(a_i^*(\mu_i, \mu_j^*, 1), a_j^*(\mu_j^*, \mu_i, \theta_j), 1) \right) \right) \\ = (1 - \mu_j^*) E_{sj1} \left(E_{\mu_i, \mu_j^*} \left(\frac{\partial u_i(a_i^*(\mu_i, \mu_j^*, 1), a_j^*(\mu_j^*, \mu_i, \theta_j), \theta_i)}{\partial \mu_j} \right) \right) \end{aligned}$$

Substituting the solutions for equilibrium actions into the payoff function reduces this to a linear equation in μ_j^* , which gives:

$$\mu_j^* = E_{sj1} \left(\frac{2E_{\mu_i}(z_j) - z_2}{z_1 - z_2} - \frac{7 + 4k_1}{8(k_1 - k_2)} \right)$$

Note that by the ‘Bayes plausibility’ condition on the distribution of posteriors induced by an information structure, as described in Kamenica and Gentzkow (2011), it must be that $E_{sj1}(E_{\mu_i}(z_j))$ is equal to the prior expectation of z_j for any $sj1$, so the following holds for \hat{V}_i regardless of player j ’s first-stage strategy:

$$\mu_j^* = \max \left\{ \frac{2E_{\hat{\mu}}(z_j) - z_2}{z_1 - z_2} - \frac{7 + 4k_1}{8(k_1 - k_2)}, 0 \right\}$$

If the parameters and the prior $\hat{\mu}$ are such that μ_j^* is positive and $\mu_j^* < \hat{\mu}$, then each player will have a partially-revealing dominant strategy in the first-stage that mixes the opponent’s posterior between 1 and μ_j^* , and the the game will have a partially-revealing equilibrium. This will be the case if:

$$\frac{1}{2} \left(\frac{7 + 4k_1}{8(k_1 - k_2)} - \frac{z_2}{z_1 - z_2} \right) < \hat{\mu} < \frac{7 + 4k_1}{8(k_1 - k_2)} - \frac{z_2}{z_1 - z_2}$$

It is straightforward to find games for which this holds; in fact, for any given prior and values of z_1 and z_2 (which determines second-stage play), one can find externalities k_1 and k_2 that lead to partially-revealing play in the first stage.

3.4.2 Cross-Effect Uncertainty

I now turn to games in which there is uncertainty over the cross-effect, x_{θ_i} . In these games players are initially uncertain about how the interaction between their own action and the

opponent's action affects their payoff. This uncertainty may be over the magnitude of the effect, or over its direction (for player i , whether player j 's action is a strategic complement or strategic substitute).

The games I consider in this section only involve uncertainty over the x parameter, and do not include externalities. Given this, I normalise y to 1 in order to simplify exposition, so payoffs have the form:

$$u_i(a_i, a_j, \theta_i) = za_i + x_{\theta_i}a_ia_j - a_i^2$$

I begin with a result that fleshes out the assumption of interior equilibrium actions for this type of games by noting what it implies about the parameters x_1 and x_2 .

Lemma 4. *If a quadratic game with $y_1 = y_2 = 1$, $z_1 = z_2$, $x_1 > x_2$, and no externalities has interior equilibrium actions for every second-stage game, then $x_1 < 1 + x_2$.*

Proof. Player j 's expected action in a game of this type is:

$$E_{\mu_i}(a_j^*(\mu, \theta_j)) = \frac{z(1 + E_{\mu_i}(x_{\theta_j}))}{1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i})}$$

This gives:

$$a_i^*(\mu, \theta_i) = \frac{z}{1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i})} \left(1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i}) + x_{\theta_i}(1 + E_{\mu_i}(x_{\theta_j})) \right)$$

It follows that if equilibrium second-stage actions are always interior, the following must hold for any posteriors μ :

$$1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i}) + x_2(1 + E_{\mu_i}(x_{\theta_j})) > 0$$

For $x_1, x_2 > 0$ this always holds as $E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i}) < 1$ by Assumption 3. For $x_1, x_2 < 0$ the minimum value of the left-hand side is $1 + x_2 < 0$, so the inequality always holds in this case as well. For $x_1 > 0$, $x_2 < 0$, the left-hand side is minimized when $E_{\mu_i}(x_{\theta_j}) = E_{\mu_j}(x_{\theta_i}) = x_1$, which gives:

$$\begin{aligned} 1 - x_1^2 + x_2(1 + x_1) &> 0 \\ 1 + x_2 &> x_1 \end{aligned}$$

□

The next result deals with uncertainty over the magnitude of the cross-effect: when actions are known to be strategic complements and the only uncertainty is over the magnitude of the complementarity, full revelation is the unique equilibrium.

Proposition 5. *For a quadratic game with $y_1 = y_2$, $z_1 = z_2$, and $x_1 > x_2 > 0$, a fully-revealing signal strictly dominates all first-stage strategies that are not fully-revealing, and consequently payoff types are revealed in any equilibrium.*

Proof. Substituting the expression for a_i^* noted in the proof of Lemma 4 into the own-type effect gives:

$$I_i(\mu_i, \mu_j) = \frac{2z^2(x_1 - x_2)^2 E_{\mu_i}(x_{\theta_j})(1 + E_{\mu_i}(x_{\theta_j}))}{(1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i}))^2} \left(1 + (x_1 + x_2) \left(\frac{1 + E_{\mu_i}(x_{\theta_j})}{1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i})} \right) \right) > 0$$

Similarly, the equilibrium-action effect is:

$$D_i(\mu_i, \mu_j) = \frac{2z^2(x_1 - x_2)^2 E_{\mu_i}(x_{\theta_j})^2 (1 + E_{\mu_i}(x_{\theta_j}))}{(1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i}))^4} \left(3E_{\mu_j}(x_{\theta_i}^2)(1 + E_{\mu_i}(x_{\theta_j})) \right. \\ \left. + 2E_{\mu_j}(x_{\theta_i})(1 - E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i})) \right) > 0$$

Since both effects are positive, \bar{V}_i is strictly convex for any beliefs. It follows that revealing information is a strictly dominant first-stage strategy. \square

The logic here is somewhat similar to that of the Cournot with cost uncertainty case of Section 3.2, although there is an equilibrium-action effect component present that was absent there. When the strategic complementarity effect on player i 's payoff is larger, the benefit to player i of an increase in player j 's action rises, both directly through the increased magnitude of the cross-effect and because player i 's own equilibrium action rises. When player j believes that the effect is larger (for player i), this raises player j 's equilibrium action, since actions are strategic complements. So an increase in μ_j benefits player i through player j 's increased action, and this benefit is itself increasing in μ_j because of the increase in the conditional expectation of $\frac{\partial u_i}{\partial a_j} = x_{\theta_i} a_i$. So the own-type effect creates convexity and an incentive to reveal.

As to the equilibrium-action effect, in these games a_j^* is convex in μ_j because of the 'feedback effect' described in the discussion of 3; as player j 's expectation of x_{θ_i} rises, so does the degree to which they expect player i to respond to an increase in player j 's action, leading to higher-order effects on a_j^* . Since $\frac{\partial u_i}{\partial a_j}$ is positive with strategic complements, this convexity in a_j^* creates a positive component in the equilibrium-action effect¹².

There is no equivalent result for games in which there is uncertainty over the magnitude of a negative cross-effect (strategic substitutes). It is still the case that a higher value of x_{θ_i} implies a higher a_i^* , but with $x_{\theta_i} < 0$ $\frac{\partial u_i}{\partial a_j} = x_{\theta_i} a_i$ is decreasing in a_i^* , so the overall own-type effect has indeterminate sign - when player i is type 1 the negative impact of an increase in a_j is reduced because $x_1 > x_2$, but increased because $a_i^*(\mu, 1) > a_i^*(\mu, 2)$.

The equilibrium-action effect is similarly variable. Player j 's action is convex in μ_j as in the strategic-complements case, but since $\frac{\partial u_i}{\partial a_j} < 0$ the corresponding component is negative. Since the component based on the change in player i 's equilibrium action is always negative, the sign of the overall effect is indeterminate.

¹²The component of the equilibrium-action effect that results from changes in player i 's equilibrium action, $\frac{\partial^2 u_i(a_i^*, a_j^*, \theta_i)}{\partial a_i \partial a_j} \frac{\partial a_j^*}{\partial \mu_j} \frac{\partial a_i^*}{\partial \mu_j}$, is always positive in quadratic games.

This implies that with strategic substitutes either convexity or concavity of \bar{V}_i is possible, and that is indeed the case; it depends on the parameter values and beliefs, so no general conclusion is possible. Numerical tests suggest that no-disclosure equilibria exist (in contrast to the case of strategic complements) for many parameter values and priors, but combinations of values also exist for which only partially-revealing equilibria are possible.

What of uncertainty as to the direction of the cross-effect? The general case of $x_1 > 0 > x_2$ is similarly variable, but a more tractable subset of these games can be found by focusing on cases in which the uncertainty is purely about direction, rather than magnitude; that is, games with $x_1 = -x_2$. Note that in this case the assumption of interior equilibrium actions implies $x_1 < \frac{1}{2}$ by Lemma 4. The next result sets out conditions under which fully-revealing and fully-concealing equilibria exist in such games, and demonstrates that for some priors, both of these outcomes are equilibria, as well as at least one partially-revealing equilibrium.

Proposition 6. *For any quadratic game with $y_1 = y_2 = 1$, $z_1 = z_2$, no externalities, $x_1 = -x_2$, and interior equilibrium actions for every second-stage game, there exists $\bar{\mu} \in (0, \frac{1}{2})$ such that:*

- *There exists an equilibrium in which both players play fully-revealing first-stage strategies if and only if $\hat{\mu} \geq \bar{\mu}$*
- *There exists an equilibrium in which both players play fully-concealing strategies if and only if $\hat{\mu} \leq \frac{1}{2}$ (that is, if the prior expectation of x_θ is negative).*
- *For $\bar{\mu} < \hat{\mu} < \frac{1}{2}$, both fully-revealing and fully-concealing equilibria exist. Additionally, at least one symmetric equilibrium exists in which first-stage strategies are neither fully-revealing nor fully-concealing.*

Proof. Write $x_1 = x$, and note that Assumption 3 and interior actions together imply that $x \in (0, \frac{1}{2})$. I begin by showing that \bar{V}_i is concave if and only if $E_{\mu_i}(x_{\theta_j}) \leq 0$. Combining the direct and own-type effects gives:

$$\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2} = \frac{4z^2 x^2 (1 + E_{\mu_i}(x_{\theta_j}))}{(1 - E_{\mu_i}(x_{\theta_j}) E_{\mu_j}(x_{\theta_i}))^4} \Phi(\mu_i, \mu_j, x)$$

where:

$$\Phi(\mu_i, \mu_j, x) = E_{\mu_i}(x_{\theta_j}) \left(1 - E_{\mu_i}(x_{\theta_j}) E_{\mu_j}(x_{\theta_i}) + \frac{3}{2} x^2 E_{\mu_i}(x_{\theta_j}) (1 + E_{\mu_i}(x_{\theta_j})) \right)$$

The sign of $\frac{\partial^2 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^2}$ is clearly determined by the sign of $\Phi(\mu_i, \mu_j, x)$. The sign of $\Phi(\mu_i, \mu_j, x)$ is in turn determined by the sign of $E_{\mu_i}(x_{\theta_j})$. This is clear when $E_{\mu_i}(x_{\theta_j}) \geq 0$, as $1 -$

$E_{\mu_i}(x_{\theta_j})E_{\mu_j}(x_{\theta_i}) > 0$. When $E_{\mu_i}(x_{\theta_j}) < 0$, we have:

$$\begin{aligned}\max_{\mu_j} \Phi(\mu_i, \mu_j, x) &= E_{\mu_i}(x_{\theta_j}) \left(1 + x E_{\mu_i}(x_{\theta_j}) + \frac{3}{2} x^2 E_{\mu_i}(x_{\theta_j}) (1 + E_{\mu_i}(x_{\theta_j})) \right) \\ &\leq E_{\mu_i}(x_{\theta_j}) \left(1 + \frac{7}{8} E_{\mu_i}(x_{\theta_j}) + \frac{3}{8} E_{\mu_i}(x_{\theta_j})^2 \right) \\ &< 0\end{aligned}$$

where the second line follows from $x < \frac{1}{2}$.

When player j plays a fully-concealing first-stage strategy \bar{s}_0 , player i 's ex-ante value function is $V_i(\mu_j; \bar{s}_0) = \bar{V}_i(\mu_j, \hat{\mu})$. This is concave if and only if $E_{\hat{\mu}}(x_{\theta_j}) \leq 0$, which is equivalent to $\hat{\mu} \leq \frac{1}{2}$. It follows that both players playing a fully-concealing strategy is an equilibrium if and only if this condition holds.

When player j plays a fully-revealing first-stage strategy, \bar{s}_1 , player i 's ex-ante value function is:

$$V_i(\mu_j; \bar{s}_1) = \hat{\mu} \bar{V}_i(\mu_j, 1) + (1 - \hat{\mu}) \bar{V}_i(\mu_j, 0)$$

Note that:

$$\frac{\partial^3 \bar{V}_i(\mu_j; \mu_i)}{\partial \mu_j^3} = \frac{12z^2 x^2 E_{\mu_i}(x_{\theta_j})^2 (1 + E_{\mu_i}(x_{\theta_j}))}{(1 - E_{\mu_i}(x_{\theta_j}) E_{\mu_j}(x_{\theta_i}))^5} \left(1 - E_{\mu_i}(x_{\theta_j}) E_{\mu_j}(x_{\theta_i}) + \frac{1}{2} x^2 E_{\mu_i}(x_{\theta_j}) (1 + E_{\mu_i}(x_{\theta_j})) \right) > 0$$

The term in parentheses here can be shown to be positive by an argument parallel to that used above to show that the sign of $\Phi(\mu_i, \mu_j, x)$ is determined by the sign of $E_{\mu_i}(x_{\theta_j})$. It follows that $\frac{\partial^3 V_i(\mu_j; s_{j1})}{\partial \mu_j^3} > 0$ for any prior, so $V_i(\mu_j; s_{j1})$ is either convex, concave, or switches from concave to convex at some intermediate value of μ_j and is concave below that threshold and convex above it.

In the latter case in which V_i switches from concave to convex, it follows that (as discussed for games with quadratic externalities in Section 3.4.1) the concave closure of $V_i(\mu_j; s_{j1})$, $\hat{V}_i(\mu_j; s_{j1})$ can be written as:

$$\hat{V}_i(\mu_j; s_{j1}) = \begin{cases} V_i(\mu_j; s_{j1}) & \mu_j \leq \mu_j^* \\ \frac{1}{1 - \mu_j^*} (V_i(1; s_{j1}) + (1 - \mu_j) V_i(\mu_j^*; s_{j1})) & \mu_j \geq \mu_j^* \end{cases}$$

where μ_j^* is the solution to the following equation if a solution exists in $[0, 1]$, and is 0 otherwise:

$$\frac{\partial V_i(\mu_j^*; \bar{s}_1)}{\partial \mu_j} = \frac{V_i(1, \bar{s}_1) - V_i(\mu_j^*; \bar{s}_1)}{1 - \mu_j^*}$$

Given this structure for $\hat{V}_i(\mu_j; s_{j1})$, a best response for player i must only put positive probability on two values of μ_j , 1 and μ_j^* . Define:

$$\nu(\mu_j, \hat{\mu}, s_{j1}) = V_i(1, s_{j1}) - V_i(\mu_j; s_{j1}) - \frac{\partial V_i(\mu_j; s_{j1})}{\partial \mu_j}$$

Given the structure described above for the concave closure, a fully-revealing first-stage equilibrium will exist if and only if:

$$\nu(0, \hat{\mu}, \bar{s}_1) \geq 0$$

This condition is necessary and sufficient because it also holds whenever $V_i(\mu_j; \bar{s}_1)$ is convex and does not hold when it is concave.

The function $\nu(0, \hat{\mu}, \bar{s}_1)$ is linear in $\hat{\mu}$ by the linearity of expectations, negative for $\hat{\mu} = 0$ since $\bar{V}_i(\mu_j, 0)$ is concave, and positive for $\hat{\mu} = 1$ since $\bar{V}_i(\mu_j, 1)$ is convex. Therefore there exists $\bar{\mu} \in (0, 1)$ such that $\nu(0, \bar{\mu}, \bar{s}_1) = 0$ and a fully-revealing equilibrium exists for $\hat{\mu} \geq \bar{\mu}$.

I now show that $\bar{\mu} < \frac{1}{2}$ by showing that $\nu(0, \frac{1}{2}, \bar{s}_1) > 0$. In that case, noting that $\bar{V}_i(\mu_j; \mu_i) = E_{\mu_j}(\frac{1}{2}y_{\theta_i}a_i^*(\mu, \theta_i)^2)$ and substituting the expressions for equilibrium actions and their derivative gives

$$\begin{aligned} \nu(0, \frac{1}{2}, \bar{s}_1) = & \frac{2x^2(1+x)^2(1-4x+4x^3-4x^4)}{(1-x)^2(1+x^2)^2} \\ & + \frac{16x^5(1+x+x^2)}{(1+x^2)^3(1+x)^2(1-x^2)} - \frac{8x^4(1+2x^4)}{(1+x^2)^3(1+x)^2(1-x^2)} \end{aligned}$$

The above is positive for any $x < \frac{1}{2}$.

These results demonstrate that if $\hat{\mu} \in (\bar{\mu}, \frac{1}{2})$, both players playing \bar{s}_1 and both players playing \bar{s}_0 are equilibria. Recall that best responses in the first stage here are mixes between 1 and some posterior, say μ_p . Abusing notation slightly, write $s_{j1} = \mu_p$ to indicate that s_{j1} is an information structure that mixes between posteriors of 1 and μ_p . Now consider the function $\nu(\mu_p, \hat{\mu}, \mu_p)$ - that is, the value of the function ν at a posterior that mirrors player j 's strategy. If $\nu(\mu_p, \hat{\mu}, \mu_p) = 0$, then there is a symmetric equilibrium in which both players use an information structure that mixes between 1 and μ_p . The function $\nu(\mu_p, \hat{\mu}, \mu_p)$ is continuous in μ_p , and the results above imply that for $\hat{\mu} \in (\bar{\mu}, \frac{1}{2})$, we have $\nu(0, \hat{\mu}, 0) > 0$ (fully-revealing strategy is the best response to a fully-revealing opponent) and $\nu(1, \hat{\mu}, 1) < 0$ (V_i is concave when opponent conceals). By the intermediate value theorem, it follows that there exists a $\mu_p \in (0, 1)$ such that $\nu(\mu_p, \hat{\mu}, \mu_p) = 0$, and a partially revealing equilibrium exists. \square

This result demonstrates the range of possible outcomes for the information sharing games I have described in this paper. In particular, it shows how the straightforward dominant-strategy equilibria found in the linear cost-uncertainty oligopoly examples can give way to more complex outcomes when the scope of uncertainty varies and makes the variation in equilibrium actions and the impact of that variation on payoffs nonlinear and/or non-monotonic. This may not even require uncertainty over the cross-effect. Proposition 3 demonstrates the possibility of fully-concealing equilibria in Cournot oligopoly with uncertainty over some kinds of quadratic cost functions, but the parameter ranges that make the ex-ante value function \bar{V}_i concave in that case depend on player i 's posterior μ_i , so it is not guaranteed to be convex for all first-stage strategies of the opponent, and multiple equilibria may exist for such games as well - an interesting open question for future work on this topic.

Bibliography

- [1] Albrecht, Brian C. (2017) “Political Persuasion.” Working paper, Kennesaw State University
- [2] Amir, Rabah, Jim Y. Jin, and Michael Troege. (2019) “Robust results on the sharing of firm-specific information: Incentives and welfare effects.” *Journal of Mathematical Economics* 46.5: 855-866
- [3] Au, Pak Hung, and Keiichi Kawai. (2019) “Competitive disclosure of correlated information.” *Economic Theory*: 1-33
- [4] Aumann, Robert J., Michael Maschler, and Richard E. Stearns. (1995) *Repeated Games with Incomplete Information*, MIT Press
- [5] Bergemann, Dirk, and Stephen Morris. (2013) “Robust predictions in games with incomplete information.” *Econometrica* 81.4: 1251-1308
- [6] Bergemann, Dirk, and Stephen Morris. (2016) “Bayes correlated equilibrium and the comparison of information structures in games.” *Theoretical Economics* 11.2: 487-522
- [7] Bergemann, Dirk, and Stephen Morris. (2019) “Information design: A unified perspective.” *Journal of Economic Literature* 57.1: 44-95
- [8] Boleslavsky, Raphael, and Christopher Cotton. (2015) “Grading standards and education quality.” *American Economic Journal: Microeconomics* 7.2: 248-79
- [9] Chokler, Adi, Shlomit Hon-Snir, Moshe Kim, and Benyamin Shitovitz. (2006) “Information disadvantage in linear Cournot duopolies with differentiated products.” *International Journal of Industrial Organization* 24, no.4: 785-793
- [10] Clarke, Richard N. (1983) “Duopolists don’t wish to share information.” *Economics Letters* 11.1-2: 33-36
- [11] Fried, Dov. (1984) “Incentives for information production and disclosure in a duopolistic environment.” *The Quarterly Journal of Economics* 99.2: 367-381
- [12] Gal-Or, Esther. (1986) “Information transmission—Cournot and Bertrand equilibria.” *The Review of Economic Studies* 53.1: 85-92

- [13] Gentzkow, Matthew, and Emir Kamenica. (2016) “Competition in persuasion.” *The Review of Economic Studies* 84.1: 300-322
- [14] Gentzkow, Matthew, and Emir Kamenica. (2017) “Bayesian persuasion with multiple senders and rich signal spaces.” *Games and Economic Behavior* 104: 411-429
- [15] Kamenica, Emir, and Matthew Gentzkow. (2011) “Bayesian persuasion.” *The American Economic Review* 101, no. 6: 2590-2615.
- [16] Kamenica, Emir. (2019) “Bayesian persuasion and information design.” *Annual Review of Economics* 11: 249-272.
- [17] Koessler, Frédéric, Marie Laclau, and Tristan Tomala. (2020) “Interactive information design.” Working paper, HEC Paris
- [18] Mathevet, Laurent, Jacopo Perego, and Ina Taneva. (2020) “On information design in games.” *Journal of Political Economy*
- [19] Novshek, William, and Hugo Sonnenschein. (1982) “Fulfilled expectations Cournot duopoly with information acquisition and release.” *The Bell Journal of Economics*: 214-218.
- [20] Raith, Michael (1996) “A general model of information sharing in oligopoly.” *Journal of economic theory* 71.1: 260-288
- [21] Taneva, Ina (2019) “Information design.” *American Economic Journal: Microeconomics* 11.4: 151-85
- [22] Vaissman Guinsburg, Pedro. (2020) “Information Design and Sensitivity to Market Fundamentals.” Working paper, University of São Paulo
- [23] Vives, Xavier. (1984) “Duopoly information equilibrium: Cournot and Bertrand.” *Journal of economic theory* 34.1: 71-94
- [24] Xu, Jin. (2010) “Duopoly information sharing with differentiated products.” *Operations research letters* 38.4: 287-291